

# Advances in Iranian Linguistics

EDITED BY  
Richard K. Larson  
Sedigheh Moradi  
Vida Samiian

JOHN BENJAMINS PUBLISHING COMPANY

# ADVANCES IN IRANIAN LINGUISTICS

# CURRENT ISSUES IN LINGUISTIC THEORY

AMSTERDAM STUDIES IN THE THEORY AND HISTORY  
OF LINGUISTIC SCIENCE – Series IV

ISSN 0304-0763

General Editor

JOSEPH C. SALMONS

University of Wisconsin–Madison

jsalmons@wisc.edu

Founder & General Editor (1975-2015)

E.F.K. KOERNER

Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin

*Current Issues in Linguistic Theory* (CILT) is a theory-oriented series which welcomes contributions from scholars who have significant proposals that advance our understanding of language, its structure, its function and especially its historical development. CILT offers an outlet for meaningful contributions to current linguistic debate.

A complete list of titles in this series can be found on [benjamins.com/catalog/cilt](http://benjamins.com/catalog/cilt)

## *Editorial Board*

Claire Bowern (New Haven, Ct.)

Alexandra D'Arcy (Victoria, B.C.)

Sheila Embleton (Toronto)

Elly van Gelderen (Tempe, Ariz.)

Iván Igartua (Vitoria-Gasteiz)

John E. Joseph (Edinburgh)

Matthew Juge (San Marcos, Tex.)

Danny Law (Austin, Tex.)

Martin Maiden (Oxford)

Martha Ratliff (Detroit, Mich.)

Klaas Willems (Ghent)

## **Volume 351**

Richard K. Larson, Sedigheh Moradi and Vida Samiian (eds.)

*Advances in Iranian Linguistics*

# ADVANCES IN IRANIAN LINGUISTICS

Edited by

RICHARD K. LARSON

*Stony Brook University*

SEDIGHEH MORADI

*Stony Brook University*

VIDA SAMIIAN

*UCLA & California State University, Fresno*

JOHN BENJAMINS PUBLISHING COMPANY  
AMSTERDAM & PHILADELPHIA



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/cilt.351

**Cataloging-in-Publication Data available from Library of Congress:**  
LCCN 2020017463 (PRINT) / 2020017464 (E-BOOK)

ISBN 978 90 272 0716 6 (HB)

ISBN 978 90 272 6093 2 (E-BOOK)

© 2020 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

# Table of contents

CHAPTER 1		
Advances in Iranian linguistics: An introduction		1
<i>Sedigheh Moradi</i>		
CHAPTER 2		
Syntactic and semantic constraints on pronoun and anaphor resolution in Persian		15
<i>Elias Abdollahnejad and Dennis Ryan Storoshenko</i>		
CHAPTER 3		
A multi-dimensional approach to classification of Iran's languages		29
<i>Erik Anonby, Amos Hayes and Robert Oikle</i>		
CHAPTER 4		
The additive particle in Persian: A case of morphological homophony between syntax and pragmatics		57
<i>Jila Ghomeshi</i>		
CHAPTER 5		
The pronoun-to-agreement cycle in Iranian: Subjects do, objects don't		85
<i>Geoffrey Haig</i>		
CHAPTER 6		
The suffix that makes Persian nouns unique		107
<i>Masoud Jasbi</i>		
CHAPTER 7		
The meaning of the Persian object marker <i>rā</i> : What it is not, and what it (probably) is		119
<i>Masoud Jasbi</i>		

CHAPTER 8	
Topic agreement, experiencer constructions, and the weight of clitics	137
<i>Thomas Jügel and Pollet Samvelian</i>	
CHAPTER 9	
Another look at Persian <i>rā</i> : A single formal analysis of a multi-functional morpheme	155
<i>Simin Karimi and Ryan Walter Smith</i>	
CHAPTER 10	
The <i>Ezafe</i> construction revisited	173
<i>Richard Larson and Vida Samiiian</i>	
CHAPTER 11	
Quantitative meter in Persian folk songs and pop lyrics	237
<i>Mohsen Mahdavi Mazdeh</i>	
CHAPTER 12	
Stripping structures with negation in Persian	257
<i>Vahideh Rasekhi</i>	
CHAPTER 13	
Oblique marking and adpositional constructions in Tat: A mosaic of dialectal convergence and divergence	275
<i>Murad Suleymanov</i>	
Author index	301
Languages index	305
Subject index	307

# Advances in Iranian linguistics

## An introduction

Sedigheh Moradi

Stony Brook University

The diffusion of Iranian languages throughout a large part of the world, their deep recorded history, the remarkably intricate typological variation they show, and their extensive contact with languages belonging to different linguistic families make Iranian linguistics a fascinating and highly promising area for research in linguistic theory. Despite this remarkable scientific potential, however, and unlike the situation with nearly every other major language family, there are few collections of works focusing exclusively on languages of the Iranian family. The present volume is part of a current ongoing effort to address this gap, providing a theoretically informative and constructive venue for scholars working in Iranian linguistics. The twelve chapters of the current volume are selected from over 40 papers presented at the first North American Conference in Iranian Linguistics (NACIL1) held at Stony Brook University, April 28–30, 2017.

**Keywords:** Iranian languages, Iranian linguistics, Persian, classification of Iranian languages, diachronic and typological studies

### 1. Iranian languages as a stable diversity hub

The main focus of the current volume is on Persian. Ten of the twelve articles focus on some aspect of Persian including historical development, morphology, phonology, syntax and semantics, as well as language typology and classification of Iranian languages.<sup>1</sup> Exploring aspects of Persian is essential in formalizing a framework within which the degrees of diversity in Iranian languages can be measured. This is not, by any means, the same as considering Persian as a standard from which other

---

1. This happened despite the fact that NACIL1 brought together scholars working on a wide range of Iranian languages including Persian, Kurdish (Sorani and Kurmanji), Caspian (Tati, Talyshi, Gilaki and Mazandarani), Zazaki, Wakhi, as well as Tajik and Afghan Sign Languages, among others.



languages deviate. Instead it is similar to canonical typology as first elaborated by Corbett (2003), which explores the characteristics of the most evident data in order to establish a theoretical space from which we can calibrate the range of possibilities and variations. This theoretical space predicts the existence of a phenomenon in a language based on the data found in other languages within the same space; or it implies the historical presence of such a phenomenon and the change it has undergone over time. One well-described example of grammatical change in Iranian deals with the alignment patterns across Iranian languages (cf. Haig 2008). Most Iranian languages are canonically verb-final with a binary marking of the verb stem based on present and past tenses. The participial form *-ta* marks the past stem, while the present stem remains as the unmarked form. Originally, *-ta* used to have an active orientation for intransitive verbs, but was passive in the transitive paradigm, as in Old Persian *hamiçiya- haghmata-* “the rebels assembled”, *ima tya mana- kartam* “this is what was done by me” (Payne & Mahmoodi-Bakhtiari 2009: 439). In later stages, the passive participle was reanalyzed as an active verb in a number of different languages, e.g., Kurmanji Kurdish *ez ketim* “I fell”, *min çîrok xwend* “I read a story.” Many Iranian languages “exhibit various stages in the decay of the past tense ergative system into a nominative one, as preserved in the tenses based on the present stem. Modern Persian is typical here of the final stage, with no traces of ergativity” (Payne & Mahmoodi-Bakhtiari 2009: 439).<sup>2</sup>

A linguistic space also represents a sizable language area which is valuable for typological studies of related languages. The observed changes are motivated either by evolutionary factors internally conditioning the Iranian family of languages or enforced under the influence of neighboring languages. This accentuates the importance of studying the languages originating in the Iranian plateau both chronologically and areally, a point elaborated on in the following section.

Despite all the changes they have undergone over centuries, Iranian languages form a fairly stable linguistic family. Stability and diversity might seem like two incompatible notions, but when studied in the manner of bio-organisms, “great diversity of genetic lineages [in a population shows] a stable feature of all areas unless specific and identifiable geographical and cultural factors intervene. ... It is an evident ancient feature in long-inhabited and linguistically autonomous parts of the world; and the frequencies of certain typological features in high-diversity areas in all parts of the globe tend to converge on a common statistical profile” (Nichols 1992: 1). This common genetic statistical profile is what also constitutes a diverse linguistic space.

---

2. Haig (2008) presents a rather different analysis of the roots of split ergativity in Iranian.

As a stable hub, the Iranian family of languages provide an excellent laboratory for studying linguistic variation within the Indo-European family. This gains utmost importance due to the archaeological findings relating the eastward and westward movement of Indo-Iranian people to the linguistic spread of Indo-European languages. In a recent study in *Science*, Narasimhan et al. (2019) build on earlier work showing “massive population movement from the Eurasian Steppe into Europe early in the third millennium BCE, likely spreading Indo-European languages” to reveal “a parallel series of events leading to the spread of Steppe ancestry to South Asia, thereby documenting movements of people that were likely conduits for the spread of Indo-European languages” (Narasimhan et al. 2019). They further conclude that the similarities between the Steppe ancestry in South Asia and the Bronze Age Eastern Europe are likely to be the reason for “the unique features shared between Indo-Iranian and Balto-Slavic languages” (Narasimhan et al. 2019).

## 2. Classification of Iranian languages

With an estimated 150 to 200 million native speakers, the Iranian languages<sup>3</sup> are spread across a vast region from West Asia (Central Turkey, Syria and Iraq) to Central Asia (Tajikistan, Afghanistan, Pakistan and Pamir) and further to the east (western part of Chinese Turkestan). To the North, they go as far as the central Caucasus (Ossetic) and North West Tajikistan (Yaghnobi), and to the South, Kumzari speakers reside in Oman across the Persian Gulf.

There is no agreed upon number of (Modern) Iranian languages and the distinction between variety and language remains contested in some cases, but some estimates claim as many as 86 languages (Eberhard et al. 2019). Persian is the language with the highest number of speakers and includes three major varieties: Iranian Persian (‘iPersian’) is the official language of Iran and the *lingua franca* serving education, media and government. Tajiki or ‘tPersian’ is another recognized variety, serving as the official language of Tajikistan; and Dari or Afghan Persian (‘aPersian’), spoken in Afghanistan, serves as the official language of Afghanistan along with Pashto, another Eastern Iranian language. Modern Iranian languages have evolved from their ‘Middle Iranian’ antecedents (Middle Persian, Parthian, Sogdian, Bactrian, Saka, etc.) which had their precursor ‘Old Iranian’ in Antiquity (Old Persian, Avestan, etc.).

In terms of linguistic genealogy, Iranian languages constitute the western group of the Indo-Iranian branch of the Indo-European language family. While

---

3. ‘Iranic’ is another term suggested to refer to this branch of languages as the anthropological name for the linguistic family (Perry 1998).



Map 1. A Geographic overview of Iranian languages

historically Iranian languages are aligned with the Indo-European, and in particular Indic languages, from the areal perspective they have long interacted with the adjacent Turkic and Semitic languages. Additionally, Iranian languages have a well-documented written history going back over 2,500 years to Old Persian and Avestan (6–3rd century BCE) and to Middle Persian and other Middle Iranian languages (3rd century BCE–7th century CE). Given their history and typological diversity, Iranian languages constitute an important source for linguistic exploration and analysis.

Traditionally, Iranian languages are classified in terms of genealogical and geographic subgroups. Windfuhr (2009) recognizes at least four major groups. The **Northwestern** group includes Kurdish, Gorani, Hawrami, Zazaki, Laki, Balochi, Talyshi, Tati, Vafsi, Sangesari, Semnani and the Caspian languages. The **Southwest** Iranian group includes the varieties of Persian: iPersian, Tajik/tPersian, Dari/aPersian, Tat in SE Caucasus, Lori, and Bakhtiari. The **Southeast** group includes Parachi andOrmuri; North or **Northeast** Iranian includes Ossetic and Yagnobi. **East Iranian**, also classified as Northeast, includes Pashto, the Pamir languages Shughni, Yazghulami, Wanji, Ishkashimi, Wakhi, Yidgha and Munji. This list simply includes some of the languages studied in recent fieldwork.

Paul (1998, 2003), in his studies of Zazaki and Balochi, emphasizes the importance of inter-dialectal borrowing throughout different historical stages, which has blurred the definitive divide between the traditional subgroups. Furthermore, he notes that the SW/NW distinction is not always geographically accurate, e.g., the NW Balochi is spoken in the SE, while the SW Tati is spoken in the NW. He proposes that since the NW/SW distinction is not clear-cut, we should rather explain the gradual changes in these languages in terms of gradation, “with each language attributed a position on a scale ranging from the ‘most northwestern’ to the ‘most southwestern’” (Paul 1998: 164). The following stratum classifies the Iranian languages between two extremes where on the “northwesternmost” end of the scale reside Talyshi, Zazaki and Gurani and on the other end there is Persian (Paul 2003: 61):

Persian < Balochi, Kurdish < Central Dialects < Caspian Languages < Semnani < Talyshi < Zazaki < Gurani

Aside from gradual internal changes that motivate a stratum-based classification of Iranian languages, there are also external changes, areal changes or changes due to contact with other language families that justify an areal classification as proposed by Stilo (2005, 2006, 2009). Working on a wide range of grammatical phenomena, Stilo demonstrates that some patterns that are generally considered ‘inconsistent’ can be accounted for as a natural result of areal influence. Stilo 2005

starts his discussion with a general claim: “Languages that do not meet the expectations of implicational universals are often found between two languages, groups of languages or language areas that are more or less opposite to each other” (Stilo 2005: 38). He further demonstrates that, within an areal typology, “Iranian languages are sandwiched” between typical VO languages (Arabic and Mediterranean languages) and typical OV languages (Turkic, North Caucasian and Indic). In this way, “Iranian languages act as a transition or buffer zone that represents a hybridization” of the opposite patterns of their neighbors (Stilo 2005: 38). He classifies this buffer zone as follows: (1) the central area (e.g., Vafsi), (2) the extreme west (e.g., Central Kurdish), and finally (3) the extreme east (e.g., Meime’i) (Stilo 2006: 313).

Given the multiple perspectives on diversity, typology, and classification of Iranian languages, the chapter by Anonby, Hayes and Oikle makes a significant contribution by presenting a novel multi-dimensional approach for classifying the languages of Iran. To overcome limitations of existing two-dimensional models of language classification, they propose a three-dimensional ‘language relation web’ based on a force-directed graph visualization as an alternative and more adequate model for expressing connections between language varieties. This architecture allows for differentiating and representing multiple types of linkages: shared genealogical inheritance, structural similarity through contact, and association through ethnic identification. The resulting model provides new insights into the classification of Iran’s languages and raises questions and prospects for the broader classification process.

### 3. Iranian linguistics

#### 3.1 Diachronic and typological studies

Descriptive studies on Persian and its historical evolution go back to the turn of the last century. Interest in Iranian languages led to some significant works by European and Soviet linguists in the descriptive tradition on Old Persian (Kent 1953) and Middle Persian (Nyberg 1923 and Rastorgueva 1964 among others). Today this tradition continues with works by Skjærvø (2006) and others. Besides studies on the historical grammar of Persian (e.g., Johnson 1917), Sadeghi (1978), Khanlari (1979), and Abolghasemi (1996) have written volumes on the historical evolution of Persian. Over the years, as Iranian languages have lost most of their morphology and have developed an intricate syntax, their morphosyntactic agreement systems have attracted many linguists. In the current volume, Jügel and Samvelian address the claim that enclitic pronouns shifted to verbal agreement

markers via topic agreement. This means that hanging topics resumed by enclitic pronouns are reanalyzed as subjects cross-indexed by agreement markers. The authors suggest a bridging context for the reanalysis of topic agreement as verbal agreement by assuming that verbal endings (the inherited agreement markers) and enclitic pronouns represent the same degree or weight of encoding. Moreover, they compare the historical findings with the cross-reference patterns found in New Persian experiencer constructions. These constructions show a similar development and provide evidence that the relation of experiencer and cross-indexing enclitic pronoun qualifies as agreement.

In addition to the diachronic and descriptive studies of Persian, comprehensive studies of individual Iranian languages and language groups have emerged in publication and dissertations (Ludwig Paul on Zazaki, and Serge Axenov on Balochi, among others). Ongoing field work and research on many of these languages, such as Balochi, continues under the supervision and expertise of linguists such as Carina Jahani, Agnes Korn, and Donald Stilo for Vafsi and Caspian languages, to name a few. Other significant collections of works on Iranian languages include the six-volume *Osnovy Iranskogo Jazykoznanija* (1979–1997) edited by Vera S. Rastorgueva, and the three-volume *Iranskie Jazyki* (1997–2000). Finally, Windfuhr's 2009 volume *Iranian Languages* presents descriptions of 16 languages representing center and outer circle of the Iranian branch.

Work on Iranian comparative typology also began early on but has taken renewed vigor with a growing group of linguists. In Iran, Dabir-Moghaddam's work on typology has primarily focused on word order parameters and other salient grammatical features (Dabir-Moghaddam 2001, 2006, 2013). Another major contribution is Haig's in-depth analysis of alignment change with data from more than 20 Iranian languages (Haig 2008). Along the same lines, his chapter in the current volume traces the evidence for cyclic grammatical change across Western Iranian. Iranian languages provide a natural historical laboratory for exploring these processes because of the long history of attestation (more than 2,000 years), and because the paradigm of subject-indexing and object-indexing pronouns is phonologically largely identical; different outcomes cannot therefore be assigned to the differences in phonological forms. The results lend support to the view that the grammaticalization of agreement of subjects is a fundamentally different process from that of object agreement. Haig further argues that despite early evidence of superficial grammaticalization (cliticization), object pronouns do not achieve full agreement status, while subject pronouns, though less archaic than the corresponding object pronouns, may do so. He thus concludes that these differences cannot be predicted by traditional grammaticalization-based accounts of the emergence of agreement, nor Minimalism-based accounts.

### 3.2 Modern studies

Bateni's 1969 description of iPersian ushered in a new era of structuralist linguistic studies inside Iran (Taleghani 2009: 279). Also presented within the same framework are Lambton (1953) and Lazard (1992). Lazar's *Grammar of Contemporary Persian* has been republished and revised many times and continues to serve as an essential resource for many linguists working on Persian.

Theoretical work on Iranian languages in the generative framework began in the 1970s, primarily focused on iPersian, beginning with John Moynes's dissertation (1970) and expanded over the years. In the *Oxford Handbook of Persian Linguistics* (Sedighi & Shabani-Jadidi 2018: Part 3, Chapter 1), Simin Karimi gives an overview of generative approaches in Persian syntax, focusing on current frameworks, in particular Minimalism. In the next chapter, Jila Ghomeshi treats 'other approaches' covering descriptive, theory neutral, functionalist and cognitivist approaches, among others.

Samiian (1983) was the first Iranian linguist to give an analysis of noun phrase and *Ezafe* construction in Persian which is also a distinguishing grammatical feature of many of the Iranian languages (Taleghani 2009: 281). In their chapter in this volume, Larson and Samiian revisit the *Ezafe* construction, addressing the nature, distribution and function of the *Ezafe* morpheme. They first review the main semantic, morphological, and syntactic analyses advanced in the wide literature on the subject. They argue that the syntactic account of *Ezafe* is the most promising, both in its empirical reach, and explanatory power. Looking at an exhaustive range of data from iPersian and other Iranian languages, they note that *Ezafe* occurs between nominal elements in the NP, AP, PP, and QPs. Following case theory (Chomsky 1981), they propose that *Ezafe* satisfies a licensing requirement in the following phrase, similar to 'of' in English. They then consider in detail the implications of this theory for the occurrence of *Ezafe* before PPs in iPersian and before finite and nonfinite complement clauses in iPersian and Kurdish. Finally, they examine the occurrence of *Ezafe* in Zazaki 'double *Ezafe* constructions' and in Caspian languages showing the so-called 'Reverse *Ezafe* construction' in light of the case-based analysis.

Following the theoretical turn in Iranian linguistics in the 1980s and 1990s, many linguists have devoted their scholarship to the development of the different grammatical aspects of these languages. Ghomeshi's (1997) work on projection and inflection in Persian started her career as an influential figure in Persian morphosyntactic studies. She contributes to the present volume by analyzing the syntactic and semantic properties of the additive marker *-am* in Persian. Through a detailed exploration of the additive marker, she proposes that some instances of homophony are illusory if we take morphological levels into account. Ghomeshi

shows that *-am* exhibits positional variability, is polysemous in meaning, and does not always contribute content affecting the truth conditions of the sentence. On the basis of this evidence, she classifies *-am* as a pragmatic particle. Noting that *-am* is homophonous with both the first person singular agreement suffix and the first person pronominal enclitic, Ghomeshi proposes that it is precisely because additive *-am* is a pragmatic particle that it is distinguished from the inflectional morphemes that it resembles in form. Drawing parallels with the phonological resemblance between the plural marker and the accusative case marker in Korean, she suggests that working out the puzzle of morphological homophony reveals significant insights into the structure of grammar. Thus, the chapter argues for three levels at which morphemes can be classified: derivation, inflection, pragmatics, and suggests that cross-level homophony is not accidental, but the frequency of use at one level predisposes a particular form to be used at another level. This ultimately gives a language its morphological ‘flavor’.

Three chapters on the extensively studied particle *-rā* provide an excellent illustration of the range of theoretical frameworks used in current scholarship. In Chapter 9, Karimi and Smith present a single formal analysis of this multi-functional morpheme. They discuss several cases in which *-rā* may appear on DPs that are not direct objects, contrary to former accounts that typically consider *-rā* a differential object marker. Building on insights from dependent case theory, they develop an analysis in which *-rā* is the realization of accusative case, treated as a dependent case assigned in syntax, as well as a specificity feature.

Another perspective is presented by Jasbi (Chapter 7), who starts by showing that *rā* is not an exclusive marker of specific or definite referents. Instead, *rā*’s core contribution, he argues, is old or presupposed information with an existence implication. A marked object such as *sandali-ro* (“chair”-*rā*) implies that there are one or more mutually known chairs in the conversation. This account captures several novel empirical observations on the distribution of *rā* such as the optional presence of *rā* on proper names in some contexts. Finally, Jasbi provides a formal and compositional analysis of simple Persian sentences with definite and indefinite objects.

Another, third, chapter (Chapter 13) dealing with *-rā* by Suleymanov presents a more extensive description of *-rā* as an Oblique marker, with a series of related functions in Tat, a less studied Iranian language spoken in the Caucasus, Azerbaijan and Dagestan. In Suleymanov’s article the various functions of the *-rā* clitic are explored in detail whether inherited or introduced through language contact, including its role in forming new types of adpositional constructions. The author uses oral and written Tat corpora to compare dialects of Tat from a specific morphological point of view. Two groups of Tat dialects are explored in detail, the Judeo-Tat varieties spoken in Dagestan and northern Azerbaijan and several Muslim Tat varieties of Azerbaijan.



Three of the four remaining chapters of this volume also deal with syntax and syntax-semantics interface in Persian. In another semantic adventure (Chapter 6), Jasbi takes up the task of analyzing the colloquial nominal definite marker *-e*. While it is widely acknowledged that Persian has no dedicated marker of definiteness, the nominal suffix *-e* has been analyzed as a colloquial definiteness marker. Here he shows that while *-e* can mark bare nominals to ensure a definite interpretation, it can also appear on indefinite nouns preceded by the indefinite determiner *ye*. He further shows that indefinites marked by *-e* are scopally inert. To unify the effect of *-e* on definites and indefinites, Jasbi proposes that *-e* introduces a uniqueness implication on the nominal it modifies. More specifically, *N-e* denotes a singleton set of objects. On a bare nominal, this uniqueness implication ensures a definite interpretation. On an indefinite, it restricts the domain of quantification to a singleton, making the indefinite scopally inert.

In Chapter 2, Abdollahnejad and Storoshenko investigate the mechanism of reference resolution in Persian for the colloquial pronoun *un* “(s)he” and the anaphor *xod-eš* “self- 3SG”. The analysis provides evidence in support of the “multiple constraints” framework for reference resolution, while expanding the relatively scant analyses of Persian anaphora and reference resolution. In identifying the differences in sensitivity to constraints, they argue that *xod-eš* and *un* are fundamentally different in their binding behavior. With respect to the semantic constraints, they observe that *un* is more sensitive to a perceiver bias than *xod-eš*. They further reinforce the bound nature of *xod-eš* in demonstrating that it requires a c-commanding antecedent. This, along with the observed tolerance for (syntactic) binding both within and beyond the local clause, motivates a comparison with the Korean *caki* and Chinese *ziji* “self”, both of which share the same properties.

Rasekhi (Chapter 12) introduces two novel Stripping constructions from Persian which occur with negation: Polarity Stripping and Negative Stripping. Both of these constructions involve clausal coordination, and in the second coordinate, the entire clause, except for a constituent, is elided under identity with corresponding parts of the first coordinate. This type of construction, she argues, involves TP ellipsis, which is licensed by the Polarity head that hosts the negative marker. She also studies Pseudo-stripping, which, despite being similar to the previous constructions, does not involve ellipsis, contrary to what has been claimed in the literature (Kolokonte 2008), and is rather derived via movement.

A major force behind all the theoretical research on Persian has been its significant literary history. Bahar’s *Stylistics* (1990), for instance, was one of the first studies of the stages in the development of Modern Persian which was mainly based on rhetorical and poetical aspects of the language. Following in that tradition, Mahdavi-Mazdeh (Chapter 11) explores quantitative meter in Persian folk

songs and pop song lyrics: the metrical system used in Persian folk songs and pop song lyrics is quantitative and follows the same general principles of Classical Persian metrics. He proposes that the apparent differences observed between the two systems originate primarily from the availability of a process of optional vowel shortening in the scansion of lines that are composed in colloquial Persian. In fact, it is mainly the phonological differences between the colloquial and formal registers of Persian, rather than purely metrical differences, that result in the split observed between these two poetic traditions. In addition to introducing optional vowel shortening, Mahdavi-Mazdeh identifies in these songs several minor deviations from the requirements of Classical Persian metrics, showing that these deviations are also systematic and can in fact be helpful in gaining a deeper understanding of Persian metrics.

These brief descriptions suggest the breadth, depth and excitement of the scholarship currently being pursued by researchers in Iranian linguistics across a wide range of languages, frameworks and theoretical orientations. These works either challenge or extend aspects of recent syntactic theory, broaden the scope of current studies on historical change or offer a new semantic approach to address a much-studied area of Persian syntax. We have every expectation for continued growth in the coming years. The community of linguists who participated in NACIL 1 reconvened for the second North American Conference in Iranian Linguistics in 2019 at the University of Arizona. Plans are currently underway for NACIL 3, to be held at the University of California, Los Angeles (UCLA) in 2021.

Finally, the editors wish to express appreciation to those who have brought this volume to life. First and foremost, we thank the chapter authors. We also thank the NACIL1 organizers and participants and the Department of Linguistics at Stony Brook University. Specifically, we would like to thank Mark Aronoff, Vahideh Rasekhi, Lori Repetti and Nazila Shafiei for all their help and support. We thank several anonymous John Benjamins reviewers suggesting many revisions which have helped improve content and organization. This conference and the resulting volume would not have been possible without the support of the Roshan Cultural Heritage Institute (RCHI), whose mission is to preserve and promote interest and research in Persian language and culture. A grant from RCHI made it possible to hold NACIL1. We wish to thank the Roshan Cultural Heritage Institute and its Founder and Chair, Dr. Elahe Omidyar Mir-Jalali, for their generous support of this conference and Iranian Linguistics over the years.

## References

- Abolghasemi, Mohsen. 1996. *Historical grammar of Persian language*. Tehran: Samt.
- Bahar, Mohammad-Taqi. 1990. *Stylistics*. Tehran: Amir Kabir.
- Batani, Mohammad-Reza. 1969. *A description of grammatical structure of Persian*. Tehran: Amir Kabir.
- Chomsky, Noam. 1981. A note on non-control PRO. *Journal of Linguistic Research* 1(4). 1–11.
- Corbett, Greville. 2003. Agreement: Canonical instances and the extent of the phenomenon. In Booi, Geert E., Janet DeCesaris, Angela Ralli, and Sergio Scalise (eds.), *Topics in morphology: Selected papers from the Third Mediterranean Morphology Meeting*, 109–128. Barcelona: Universitat Pompeu Fabra.
- Dabir-Moghaddam, Mohammad. 2013. *Typology of Iranian languages*. Tehran: SAMT.
- Dabir-Moghaddam, Mohammad. 2006. Internal and external forces in typology: Evidence from Iranian languages. *Journal of Universal Language* 7(1). 29–47.  
<https://doi.org/10.22425/jul.2006.7.1.29>
- Dabir-Moghaddam, Mohammad. 2001. Word order typology of Iranian languages. *The Journal of Humanities of the Islamic Republic of Iran* 8(2). 17–23.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the world*. 22nd edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>
- Ghomeshi, Jila. 1997. Non-projecting nouns and the ezafe: Construction in Persian. *Natural Language and Linguistic Theory* 15(4). 729–788. <https://doi.org/10.1023/A:1005886709040>
- Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A construction grammar approach*. Berlin: Walter de Gruyter. <https://doi.org/10.1515/9783110198614>
- Iranskie lazyki. (3 vols): Yugo-Zapadnye Iranskie lazyki. 1997; Severo-Zapadnye Iranskie lazyki. 1999; Vostochnye Iranskie lazyki. 2000.
- Johnson, Edwin Lee. 1917. *Historical grammar of the ancient Persian language* (Vol. 8). New York: American Book Company.
- Kent, Ronald G. 1953. *Old Persian: Grammar. Texts. Lexicon*. New Haven: American Oriental Society.
- Khanlari, Parviz N. 1979. *A history of the Persian language* (Vol. 2). Tehran: Sterling Publishers.
- Kolokonte, Marina. 2008. Bare argument ellipsis and information structure. Newcastle: Newcastle University dissertation.
- Lambton, Ann K. S. 1953. *Persian grammar*. Cambridge: Cambridge University Press.
- Lazard, Gilbert. 1992. *A Grammar of Contemporary Persian*. Vol. 106. Costa Mesa, CA: Mazda Publishers.
- Moyné, John Abel. 1970. *The Structure of Verbal Construction in Persian*. Cambridge: Harvard University dissertation.
- Narasimhan, Vagheesh M., Nick Patterson, Priya Moorjani, Nadin Rohland, Rebecca Bernardos, Swapan Mallick, et al. 2019. The formation of human populations in South and Central Asia. *Science* 365, eaat7487. <https://doi.org/10.1126/science.aat7487>
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226580593.001.0001>
- Nyberg, Henrik Samuel. 1923. The Pahlavi documents from Avromân. In *Le Monde Oriental* 17(2/3). 182–230.

- Paul, Ludwig. 2003. The position of Balochi among Western Iranian languages: The verbal system. In Carina Jahani and Agnes Korn (eds.), *The Baloch and their neighbors: Ethnic and linguistic contact in Balochistan in historical and modern times*, 61–71. Wiesbaden: Reichert Verlag.
- Paul, Ludwig. 1998. The position of Zazaki among West Iranian languages. In Nicholas Sims-Williams (ed.), *Proceedings of the Third European Conference of Iranian Studies held in Cambridge. Part I: Old and Middle Iranian studies*, 163–177. Wiesbaden: Dr. Ludwig Reichert.
- Payne, John R., and Behrooz Mahmoodi-Bakhtiari. 2009. Iranian languages. In Bernard Comrie (ed.), *The world's major languages*, 453–460. London: Routledge.
- Perry, John R. 1998. A review of the “Encyclopaedia Iranica”. *Iranian Studies* 31(3/4). 517–525. <https://doi.org/10.1080/00210869808701929>
- Rastorgueva, Vera S. 1964. *Opyt sravnitel'nogo izučenija tadžikskich govorov*. Moskva: Nauka.
- Rastorgueva, Vera S. (ed.). 1979–1997. *Osnovy Iranskogo jazykoznanija* (6 vols). Moskva: Nauka.
- Sadeghi, Ali Ashraf. 1978. *Development of Persian language*. Tehran: Daneshgah-e Azad.
- Samiian, Vida. 1983. Structure of phrasal categories in Persian: An X-bar analysis. Los Angeles, CA: UCLA dissertation.
- Sedighi, Anousha, and Pouneh Shabani-Jadidi (eds.). 2018. *The Oxford handbook of Persian linguistics*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198736745.001.0001>
- Skjærvø, Prods Oktor. 2006. Iran vi. Iranian languages and scripts. *Encyclopædia Iranica*, XIII/4. 344–377, available online at <http://www.iranicaonline.org/articles/iran-vi-iranian-languages-and-scripts> (accessed on 30 December 2012).
- Stilo, Donald. 2009. An introduction to the Araxes-Iran linguistic area. Manuscript. Leipzig: *Max Planck Institute for Evolutionary Anthropology*.
- Stilo, Donald. 2006. Circumpositions as an areal response: The case study of the Iranian zone. In Lars Johanson and Christiane Bulut (eds.), *Turkic-Iranian contact areas: Historical and linguistic aspects*, 310–333. Wiesbaden: Harrassowitz Verlag.
- Stilo, Donald. 2005. Iranian as buffer zone between the universal typologies of Turkic and Semitic. In Éva Ágnes Csató, Bo Isaksson and Carina Jahani (eds.), *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic*, 35–63. London; New York: Routledge.
- Taleghani, Azita H. 2009. Persian linguistics in the twentieth century. *Iranian Studies* 42, 2. 275–284. <https://doi.org/10.1080/00210860902765000>
- Windfuhr, Gernot. (ed.). 2009. *The Iranian languages*. Florence: Routledge.



# Syntactic and semantic constraints on pronoun and anaphor resolution in Persian

Elias Abdollahnejad and Dennis Ryan Storoshenko

University of Calgary

This chapter investigates the mechanism of reference resolution for the colloquial pronoun *un* “(s)he” and the anaphor *xod-eš* “self-3SG” in Persian. The present analysis serves as another piece of evidence in support of the ‘multiple constraints’ framework for reference resolution advanced in Kaiser (2003) and Kaiser et al. (2009), while also contributing to the relatively scant analyses of Persian reference resolution. In identifying the differences in sensitivity to constraints, though, we argue that *xod-eš* and *un* are fundamentally different in their binding behaviour. Regarding the semantic constraints, we observed that *un* is more sensitive to a perceiver bias than *xod-eš*. We further reinforce the bound nature of *xod-eš* with the observation that it requires a c-commanding antecedent. This, along with the observed tolerance for (syntactic) binding both within and beyond the local clause, motivates its comparison with the Korean *caki* and Chinese *ziji* “self”, both of which share these properties.

**Keywords:** binding, pronouns, anaphors, logophoricity, multiple constraints framework

## 1. Introduction

This chapter is concerned with the mechanism of reference resolution for two forms in Persian: the colloquial pronoun *un* “(s)he” and the anaphor *xod-eš* “self-3SG” as shown in (1):

- (1) *Sohrāb<sub>i</sub> be Āraš<sub>j</sub> goft [ke Minā<sub>k</sub> hatman bā un<sub>i/j/\*k</sub> / xod-eš<sub>i/j/#k</sub>*  
Sohrab to Arash said that Mina certainly with (s)he self-3SG  
*tamās mi-gire].*

contact DUR-get

“Sohrab<sub>i</sub> said to Arash<sub>j</sub> that Mina will certainly contact him<sub>i/j/\*k</sub> / self<sub>i/j/#k</sub>.”

In this example, the pronoun shows a clear Condition B effect, resisting local binding. However, the behaviour of *xod-eš* is more unexpected. Although a local subject is generally the most likely antecedent for the anaphor, the lower predicate's semantics precludes a reflexive reading, as *tamās gereftan* "to contact" is not generally a reflexive action. Conversely, a predicate such as *dust dāštan* "to like" would allow equally ambiguous local and long distance binding. Like *un* then, *xod-eš* can also take either matrix argument as its antecedent, though such interpretations add emphasis, adding a sense of contrastive focus similar to the English *he himself*. In this chapter, we argue that while both forms can appear in overlapping environments, and are subject to some of the same constraints, the reference resolution mechanisms for *un* and *xod-eš* are different. Specifically, we claim that *un* functions as a 'standard' co-referential pronoun, drawing its reference from context alone, while *xod-eš* shows some hallmarks of a semantically bound anaphor. Crucially, as is already clear from the data in (1), the contrast between these two forms is not as simple as a straightforward Condition A vs. Condition B effect. Rather, we will compare the Persian data with examples in extensively-studied East Asian languages. Our analysis serves as another piece of evidence in support of the 'multiple constraints' framework for reference resolution advanced in Kaiser (2003) and Kaiser et al. (2009), while also contributing to the relatively scant analyses of Persian reference resolution, which largely use English as a point of departure.

The rest of this chapter has the following structure. In § 2, we present the relevant Persian facts, including a summary of two existing accounts, culminating with a fuller comparison of the distributions of the two forms, *un* and *xod-eš*. Section 3 presents the multiple constraints framework in more detail, using illustrations from English and Korean. The latter data will be especially useful in providing a set of tests to determine whether *xod-eš* can be more accurately classified as a long-distance anaphor akin to the Korean *caki* or Mandarin *ziji*. This cross-linguistic data comparison, drawing on recent analyses of these East Asian languages, comprises the bulk of § 4. Section 5 closes the chapter with a summary and prospects for future research.

## 2. Core Persian binding data

One of (if not the) earliest studies in Persian binding is that presented in Moyne (1971). In this chapter, a detailed analysis of the properties of the anaphor *xod-eš* is presented along with the related monomorphemic *xod* "self" and possessive form *xod-e un*, which translates roughly as "his/her own self" in Moyne's analysis. Interestingly, and contra later work, Moyne (1971: 145) claims that *xod* is not

permitted in isolation in the colloquial language, and can only be used as part of an *Ezafe* structure with a pronoun or full nominal. However, as part of this complex structure, *xod* could be introduced directly into the derivation as an emphatic element with no obligatory sense of reflexivity. One of the most crucial facts Moyne attempts to capture is that (2) is ambiguous:

- (2) *Hušang<sub>i</sub> xod-eš<sub>ij</sub> =o did.*<sup>1</sup>  
 Hushang self-3SG =OM saw  
 a. “Hushang<sub>i</sub> saw himself<sub>i</sub>.”  
 b. “Hushang<sub>i</sub> saw him<sub>j</sub> himself<sub>j</sub>.” (Moyne 1971: 155 Example 72 & 73)

For Moyne, the difference in the readings is a result of two different transformational derivations. The ‘true’ reflexive reading (a) is the result of a Lees & Klima (1963)-style reflexivization transformation of the object *Hušang* into *xod-eš*. The non-coreferential reading (b) is taken to be the result of a derivation whose underlying object is a contra-indexed *xod-e u*, which would escape the reflexivization transformation, but then be subject to a separate cliticization rule. While the analysis makes use of derivational machinery which has long since fallen out of use, the insight here is that the two readings are to be attributed to a contrast in emphasis: the unbound reading is the result of a derivation in which the *xod* enters the derivation as an emphatic element.

The analysis of such forms (in fact, using the exact same predicate) is taken up again in Ghomeshi & Ritter (1996), who present a version of (2) with the same indexation. However, they contradict Moyne’s statement and also present a version of this sentence with *xod* as the standalone object:

- (3) *Jiân<sub>i</sub> xod<sub>i/ra</sub>-râ did.*<sup>2</sup>  
 Jian self-CASE saw  
 “Jian saw himself.” (Ghomeshi & Ritter 1996: 94 Example 26a)

Similarly, Ghomeshi & Ritter provide (4), which mirrors Moyne’s assumed underlying form for one of the readings of (2):

1. In the original text Moyne does not provide a word by word gloss of the sentence. He also uses the more formal form *xod-aš =ra* instead of *xod-eš =o*. In this chapter we use OM in the Persian example sentences for the ‘Object Marker’ particle *=râ* or its informal counterparts *=ro* or *=o*.

2. This example is presented as in the original source, i.e. Ghomeshi & Ritter (1996). They have used *-râ* and *-o* (glossed as Case) for the object marker instead of OM which we have used in our Persian examples.



- (4) *Jiân<sub>i</sub> xod<sub>j</sub>-e u<sub>ij</sub>-o did.*<sup>3</sup>  
 Jian self PRON(3SG)-CASE saw  
 “Jian saw HIM.” (Ghomeshi & Ritter 1996: 94 Example 26c)

Of course, the analysis presented in this later work is couched in the more familiar terms of the binding conditions. For Ghomeshi & Ritter, *xod-eš* is more properly analyzed as a possessive DP [<sub>DP</sub> *xod<sub>ij</sub>-eš pro<sub>ij</sub>*] in which the *pro* possessor at a rightward [Spec, DP] provides a local binder for the *self* form and serves to determine the binding properties of the larger DP as a whole. In the case of (2), their analysis is that the enclitic *-eš* provides agreement for the *pro*, making the DP itself the relevant governing category (binding domain) in which *pro* must be free. Beyond this, either the free or the bound reading is possible. Conversely, (4) does not contain the enclitic, making the entire clause the domain in which *u* (and therefore the whole DP) must be free. Crucial to this analysis is that *xod* itself must not have any  $\Phi$ -features, as these would play the same role as the enclitic in providing the key agreement portion of the accessible subject discussion. The implication here is that *xod* should follow Condition A, pronouns such as *un* follow Condition B, and the complex form *xod-eš* follows neither, able to be free or bound in a simplex clause. It is worth noting that this is exactly the opposite of the typological prediction made by Reinhart & Reuland (1993), where the typological generalization is that monomorphemic forms lacking in  $\Phi$ -features tend to have more binding possibilities than related bi-morphemic forms.

While both of these analyses pay close attention to the binding properties within a single clause, much less attention is given to the behaviour of the various forms across clauses. The local domain provides the sharpest distinctions between the various forms, but there are untested predictions about the behaviour of *xod-eš* across clauses. For Moyne, the prediction is that any instance of *xod-eš* which is not locally bound is the result of an underlyingly emphatic structure, as only locally bound instances would have been formed by the reflexivization transformation. However, binding across clauses should indeed be possible, possibly as a result of his encliticization transformation applying after pronominalization across simplex sentences. For Ghomeshi and Ritter, the prediction is essentially the same, although there is no need to posit a difference between local and long-distance bound cases. Assuming *xod-eš* is able to be bound or free within the most local clause, it would

3. This sentence is also presented as in the original source. They have used the more formal 3rd person singular pronoun *u* (glossed as “pron(3SG)”) instead of the informal *un* (glossed as “(s)he” in our examples). Moreover, the object marker *-o* in this example seems to be a typo. Since the preceding sound is a vowel (*u*), and it is the formal form of the 3rd person singular pronoun, using the formal object marker *=ra* instead of *=o* renders the sentence more acceptable for speakers. The *Ezafe* attached to *xod* is also considered irrelevant and not glossed in the original form.

be unexpected to find obligatory binding from higher clauses. To sum up the picture of *xod-eš* obtained thus far, it would be ‘essentially pronoun-like (indifferent to being bound or not), but not even subject to Condition B’ (Gomeshi & Ritter 1996). In other words, across clauses, we should expect the differences between *xod-eš* and *un* to be minimal.

We should also clarify at this point that we have so far been using the term ‘bound’ only in its syntactic sense of describing a situation of co-indexation under a c-commanding antecedent. In the next section, we will contrast this with a more restricted semantic definition of binding, which generally assumes syntactic binding as a necessary but not a sufficient condition. First, we will introduce the multiple-constraints framework.

### 3. Multiple constraints

Although syntactic information has been considered as the major determining constraint in reference resolution, semantic information, especially in cross-clausal referential relations, has also been argued as an influential factor. According to the ‘form-specific multiple-constraints framework’ (Kaiser 2003; Kaiser et al. 2009), reference resolution as a form-specific process, is determined by the interaction of multiple types of constraints (i.e. syntactic, semantic/discourse), each weighed differently in different forms and positions. Arnold (2001, as cited in Kaiser et al. 2009) points out that the thematic role of a potential antecedent can affect its likelihood of co-reference. Specifically, some researchers have pointed to the importance of semantic information in determining referential relations, e.g. preference for source of information as antecedent of reflexives (Kuno 1987) and perceiver of information as the antecedent of pronouns (Tenny 2003). Opposing the hypothesis that ‘the relative weights of syntactic and semantic constraints are the same for reflexives and pronouns’, Kaiser et al. (2009) found that both structural and semantic information influence the referential interpretation of pronouns and reflexives in English, although at different degrees of sensitivity. The contrast in question for Kaiser et al. involves the interaction between a (structural) preference for binding from subjects and a (semantic) manipulation of the source/perceiver status of the subject in (5) and (6):

(5) *Peter<sub>i</sub> told Andrew<sub>j</sub> about the picture of him<sub>i/j</sub>/himself<sub>i/j</sub> on the wall.*  
(Kaiser et al. 2009: 60 Example 8)

(6) *Peter<sub>i</sub> heard from Andrew<sub>j</sub> about the picture of him<sub>i/j</sub>/himself<sub>i/j</sub> on the wall.*  
(Kaiser et al. 2009: 60 Example 8)

The question is whether the pronoun will behave differently from the anaphor in terms of preference for selecting the subject antecedent from inside the picture noun phrases, e.g. *picture of him/himself* in (5) & (6). Their findings, using a picture matching task, are given in Figure 1. It is clear that the reflexive *himself* in this position is strongly subject oriented, regardless of the subject's thematic role. For the pronoun *him*, the situation is more nuanced: it is not the case that semantics is the only factor, though there is a clear indication that perceivers are preferred over sources as the antecedent for the pronoun. This effect is modulated by a lingering structural preference, even when the subject is a source, as in the leftmost column of Figure 1.

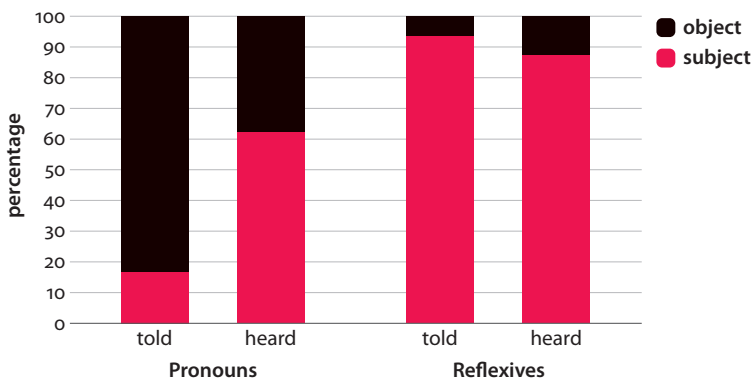


Figure 1. Kaiser et al. (2009) results on English reference resolution biases

The same effect can be seen in Persian by comparing (1), repeated below as (7) and (8):

- (7) *Sohrāb<sub>i</sub> be Āraš<sub>j</sub> goft [ke Minā<sub>k</sub> hatman bā un<sub>i/j/\*k</sub> / xod-eš<sub>i/j/#k</sub> tamās mi-gire].*  
 Sohrab to Arash said that Mina certainly with (s)he self-3SG

contact DUR-get

“Sohrab<sub>i</sub> said to Arash<sub>j</sub> that Mina<sub>k</sub> will certainly contact him<sub>i/j/\*k</sub> / self<sub>i/j/#k</sub>.”

- (8) *Sohrāb<sub>i</sub> az Āraš<sub>j</sub> šenid [ke Minā<sub>k</sub> hatman bā un<sub>i/j/\*k</sub> / xod-eš<sub>i/j/#k</sub> tamās mi-gire].*  
 Sohrab from Arash heard that Mina certainly with (s)he self-3SG

contact DUR-get

“Sohrab<sub>i</sub> heard from Arash<sub>j</sub> that Mina<sub>k</sub> will certainly contact him<sub>i/j/\*k</sub> / self<sub>i/j/#k</sub>.”

The contrast in the local clause is consistent in both examples, and exactly follows the descriptions in the previous section. At first glance, it also seems that the prediction for equivalent behaviour in long distance environments has been

met. However, we should recall that the English (5) and (6) were also technically ambiguous, but with significantly different preferences. Though it remains to be experimentally tested in a lab setting, the one author with native speaker judgments reports a similar difference in preferences for Persian. With neutral intonation, *xod-eš* in both sentences is more likely to be bound by matrix subject, i.e. source of information in (7) and perceiver in (8). However, the pronoun in both sentences is less clear, and it seems that semantics also plays a major role as the non-subject perceiver *Āraš* in (7) is the most likely antecedent for *un*. In (8), both *un* and *xod-eš* are most likely to take the subject as antecedent, but this could be due to a convergence of a semantic bias for perceivers and a structural bias for subjects, respectively. Overall, we take this as evidence that there is a distinction between the two forms, which would not be uncovered simply by enumerating their potential antecedents.

#### 4. Long distance anaphors

Anaphors which can be bound either locally or from higher clauses are quite well-known, particularly in the study of East Asian languages. Building on the evidence of the last section, we will set aside our typological bias and compare two of those monomorphemic forms, Korean *caki* “self” and Mandarin *ziji* “self” with the more complex *xod-eš*. We explore the question of whether this is a more informative line of comparison than with English. Like *xod-eš*, both of these forms can be used ambiguously for local or long-distance reference:

- (9) *John<sub>i</sub>-i [Mary<sub>j</sub>-ka caki<sub>i/j</sub>-lul salangha-n-tako] sayngkakha-n-ta.*  
 John-NOM Mary-NOM self-ACC love-PRS-COMP think-PRS-DECL  
 “John thinks that Mary loves self.” (Yoon 1989: 480 Example 5)

- (10) *Zhangsan<sub>i</sub> renwei [Lisi<sub>j</sub> zhidao [Wangwu<sub>k</sub> xihuan ziji<sub>i/j/k</sub>]]*  
 Zhangsan think Lisi know Wangwu like self  
 “Zhangsan thinks that Lisi knows Wangwu likes self”  
 (Cole et al. 1990: 1 Example (1))

The comparison to Korean here is most telling, as this form has long been argued to show the same kind of source bias interacting with a structural one, as in this example drawn from Yoon (1989):

- (11) *John<sub>i</sub>-i Mary<sub>j</sub>-eykey [caki<sub>i/j</sub>-ka am-i-lako] malha-yess-ta.*  
 John-NOM Mary-DAT self-NOM cancer-be-COMP say-PST-DECL  
 “John told Mary that self has cancer.” (Yoon 1989: 481 Example 9a)

Yoon reports that in (11), the anaphor *caki* can only take the matrix subject as its antecedent. However, in a parallel example with the verb *tul* ‘hear’, the indirect object is also a potential antecedent. Sohng (2004), reporting a similar pattern, describes this phenomenon as ‘weak subject orientation’: a violable structural preference for subject antecedents. We believe this accurately describes the state of affairs in Persian (7) and (8): *xod-eš* has an underlying preference for subject, but this may be mitigated depending on the predicate. Testing this hypothesis for Korean, Han et al. (2015) report on an eye-tracking study following Kaiser et al.’s (2009) methodology (but using sentences of the same basic form as (11)), comparing the binding properties of *caki* with a null pronoun in the same clause-initial position. Their findings are that *caki* shows a clear preference for subject antecedents upon first processing, but this can be mitigated somewhat by the predicate’s manipulation of the position (matrix subject or matrix indirect object) of an information source/perceiver. Much as in our (7) and (8), *caki* would be reported as technically ambiguous, but with a strong preference to the subject. Conversely, the null pronoun shows no structural bias, and is more clearly responding to the change in the predicate. The parallel to (7) and (8) is clear: *xod-eš* mirrors the pattern of *caki* while *un* is patterning along with the Korean null pronoun.

In light of this parallel, we now explore the possibility that *xod-eš* can be characterized as a bound variable, using a subset of the diagnostics used for *caki* in Han & Storoshenko (2012). The authors claim that beyond syntactic binding, *caki* is more accurately portrayed as being obligatorily bound by a lambda operator in the semantics. That is, the long distance reading of (9) would not be the result of co-indexation being resolved by way of an assignment function; rather the reading would result from *caki*’s co-variance with a lambda operator arising from a generalized quantifier treatment of the proper name *John*. (Barwise & Cooper 1981; Büring 2005). To start off on this line, we point out the simple fact that *xod-eš* can function as a bound variable, taking the quantifier *hær-kæsi* as an antecedent in (12):

- (12)  $hær-kæsi_i \ un_{*ij} =o \ / \ xod-eš_i =o \ / \ xod_i =rā \ dust \ dāre.^4$   
 every-body s(he) =OM self-3SG =OM self =OM friend have  
 ‘Everybody<sub>i</sub> likes him<sub>\*ij</sub> / self<sub>i/\*j</sub>.’<sup>4</sup>

Indeed, both *xod* and *xod-eš* can function as bound variables. However, echoing Moyné’s claim that there are potentially two different versions of *xod-eš*, a sufficiently emphatic context may lead to an unbound reading. The same is not true of *xod*, which is obligatorily bound as Ghomeshi & Ritter state. However, the pronoun

4. The monomorphemic *xod* is formal, so the formal object maker =*rā* attached to it, instead of =*ro* attaching to *xod-eš* and the pronoun *un*.

*un* will only provide a non-local referential reading. Extending this test for a bound variable nature over to VP ellipsis, as in (13), we have similar results:

- (13) *Sohrāb<sub>i</sub> un<sub>\*i/j</sub> =ro / xod-eš<sub>i/\*j</sub> =ro / xod<sub>i/\*j</sub> =rā dust dāre, vāli sārā<sub>j</sub>*  
 Sohrab s(he) =OM self-3SG =OM self =OM friend have but S.  
*un<sub>i/\*j</sub> =rə / xod-eš<sub>\*i/j</sub> =rə / xod<sub>\*i/j</sub> =rā dust næ-dāre.*  
 s(he) =OM self-3SG =OM self =OM friend NEG-have  
 “Sohrab<sub>i</sub> likes him-her<sub>\*i/j</sub> / self<sub>i/\*j</sub>, but Sara<sub>j</sub> doesn’t like him-her<sub>i/\*j</sub> / self<sub>\*i/j</sub>.”

The preferred reading for *xod-eš* is the sloppy reading where we find covariance, and thus semantic binding. Again though, sufficient emphasis can yield a strict reading. This is in contrast with *xod*, which only allows a sloppy bound variable reading. Conversely, *un* does not allow a sloppy bound variable reading. In addition to retaining the ability to refer to an individual outside the sentence, a sufficiently rich discourse context allows the first and (elided) second instances of *un* to behave in a way that can be described as sloppy contra-indexation. In other words, *un* in the antecedent clause may refer to the subject of the clause containing ellipsis, while the elided *un* may refer to the subject of the antecedent clause. With these examples, we can see that *xod-eš* is showing some traits of a bound variable anaphor, though the ability to escape binding is puzzling, and calls for more rigorous testing and/or corpus analysis to determine the contexts in which the strict readings can occur.

However, strict and sloppy readings are not the only diagnostics for bound variables. Anand (2006) considers logophoricity as a type of semantic variable binding in his discussion of the Mandarin anaphor *ziji*. One simple test for semantic binding arising from this is the behaviour of multiple instances of *xod-eš* within a single sentence. With no intervening potential antecedents, if *xod-eš* is indeed subject to semantic binding, then both instances should be forced to take the same antecedent.

(14) shows this to be the case:

- (14) *Sohrāb<sub>i</sub> mi-dune [ke faqat mādar-e xod-eš<sub>i</sub> hichvaqt xod-eš<sub>i</sub> =o*  
 Sohrab DUR-know that only mother-EZ self-3SG never self-3SG =OM  
*tanhā ne-mi-zāre].*  
 alone NEG-DUR-put.  
 “Sohrab<sub>i</sub> knows that only self<sub>i</sub>’s mother does not leave self<sub>i</sub> alone.”

Following Anand (2006), this obligatory co-reference, especially under an attitude predicate such as *dunestan* “to know”, is suggestive of logophoric (semantic) binding. Crucially, multiple instances of *un* within the same environment do not have this same obligatory co-reference. The observation of this forced co-reference of multiple instances of *xod-eš*, especially across a clause boundary, is especially troublesome for any analysis that maintains *xod-eš* has pronoun-like properties. Furthermore, the apparent unavailability of contra-indexation for the two instances

in (14) calls into question the idea that there may be more than one derivational route to a *xod-eš*.

A further test for logophoricity, Anand's (2006) *de re* blocking effect, is based on the distinction in (15) and (16):

- (15) *Sohrāb<sub>i</sub> fekr kard [ke Āraš<sub>j</sub> be un<sub>i</sub> gofte [ke māšin-e xod-eš<sub>ij</sub> =o dozdīd-an]].*  
 =OM stole-3PL

‘‘Sohrab<sub>i</sub> thought that Arash<sub>j</sub> has said to him<sub>i</sub> that they have stolen self<sub>ij</sub>'s car.’’

- (16) *Sohrāb<sub>i</sub> fekr kard [ke Āraš<sub>j</sub> be pedar-e un<sub>i</sub> gofte [ke māšin-e xodeš<sub>ij</sub> =o dozdīd-an]].*  
 self-3SG =OM stole-3PL

‘‘Sohrab<sub>i</sub> thought that Arash<sub>j</sub> has said to his<sub>i</sub> father that they have stolen self<sub>ij</sub>'s car.’’

The crucial distinction here lies in the positioning of the pronoun *un*, which is co-indexed with the matrix subject and *c*-commands *xod-eš* in (15) but not (16). If the expression in question, *xod-eš* in this case, is a logophor, then the presence of a syntactically binding pronoun will compete with a higher up semantic binder with the same index, rendering the binding impossible. The only possible reading is the one that avoids this competition through binding by the contra-indexed subject of the medial clause. This competition is present in (15) but not (16) where the pronoun is not in the *c*-command path between *xod-eš* and *Sohrab*. The results are not clear-cut, but in (15), the more local subject *Āraš* is the best antecedent for *xod-eš*, while the matrix subject is more likely in (16). Again, the intuitions we have gathered thus far are shaky, emerging as preferences rather than binary choices of (un)grammaticality, but the fact that *xod-eš* is sensitive to this test at all is evidence toward logophoric binding.

## 5. Conclusion

Returning to our original goals from the outset of the chapter, we do believe that Persian can provide fertile ground for the application of Kaiser et al.'s multiple constraints framework. With at least three forms to compare, *xod*, *xod-eš*, and *un*, there are many open questions around sorting out the relative weights of structural (subject orientation and basic *c*-command) constraints on the interpretation of these forms as compared to semantic constraints, be they informed by Tenny-inspired source/perceiver contrasts, or possibly the influence of logophoric environments

under attitude predicates. We have scratched the surface of this with our observation that *un* is more sensitive to a perceiver bias than *xod-eš* is, but there is clearly much more work to do in this area. In identifying these differences in sensitivity to constraints though, we have argued that *xod-eš* and *un* are fundamentally different in their binding behaviour, which may not be in line with the Ghomeshi & Ritter (1996) treatment of *xod-eš* taking its binding possibilities from a *pro* looking outside of its governing category. In essence, they describe *xod-eš* as free or bound within the minimal clause, outside the *pro*'s governing category in their account, which we have taken to predict that outside the minimal clause, *un* and *xod-eš* will have identical behaviours. As we have demonstrated, this is not the case.

The bound nature of *xod-eš* was further re-inforced with the observation that it seems to require a *c*-commanding antecedent. This, along with the observed tolerance for (syntactic) binding both within and beyond the local clause, is the motivation for our comparison with *caki* and *ziji*, both of which share these properties. While the results are not clear, the observed preferences always tend to lean toward a semantically bound variable analysis. This is not surprising given that *xod* appears to quite clearly act as a bound variable under ellipsis. Though the example with multiple instances of *xod-eš* is the clearest argument against it, we believe that Moyne's initial idea of a formal ambiguity in the derivation of *xod-eš* should not be hastily dismissed. Anand (2006) makes extensive use of this idea to account for inter-speaker variation in Mandarin, while Han & Storoshenko similarly resort to lexical ambiguity as a final cover for cases which do not fit semantic binding. Overall, the picture is that *un* and *xod-eš* are subject to different but possibly overlapping sets of constraints with different weights for each, but we are not yet convinced that a Moyne-inspired analysis, now under the guise of lexical ambiguity, should be abandoned for *xod-eš*. Such an account might explain the unexpectedly unbound readings. This of course means further research would be needed to tease apart the behaviour of multiple versions of *xod-eš*, but it could address the recurring theme that emphasis plays a role in broadening the range of interpretations.

Finally, we would be remiss in failing to point out that we have limited our discussion in this chapter only to third person singular *xod-eš*. The pronominal clitic portion of this anaphor can of course take any of 1st, 2nd, or 3rd persons, in singular or plural. This is possibly the most striking difference from the East Asian cases, as *caki* is inherently third person, while *ziji* is person-neutral, though Anand does propose what it is inherently first person for at least some speakers. Given the well-known differences between participant (1st, 2nd) versus non-participant (3rd) persons in variable binding, we expect this comparison within the *xod-clitic* paradigm would prove enlightening. As, indeed, would comparison with the undiscussed complex DP construction *xod-e-un*, which again shows all person/number combinations. All of these avenues are opened for future work.



## Funding information

Funded by: SSHRC Insight Development Grant.

Award ID: 430-2016-00128.

Award recipient: Storoshenko.

Statement: This research is funded by SSHRC Insight Development Grant 430-2016-00128 to Storoshenko. All errors are our own.

## Acknowledgements

We would like to thank the audience at the NACIL-1 conference for their feedback and comments on this work. We would like to especially thank Jila Ghomeshi for her helpful comments. Additionally, thanks to the graduate students and faculty of the Linguistics Program at the University of Calgary for their feedback during the development of this chapter.

## References

- Anand, Pranav. 2006. De de se. Cambridge, MA: MIT dissertation.
- Barwise, Jon & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4 (2). 159–219. <https://doi.org/10.1007/BF00350139>
- Büring, Daniel. 2005. *Binding theory*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511802669>
- Cole, Peter, Gabriella Hermon & Li-May Sung. 1990. Principles and parameters of long-distance reflexives. *Linguistic Inquiry* 21. 1–22.
- Ghomeshi, Jila & Elizabeth Ritter. 1996. Binding, possessives, and the structure of DP. In Kiyomi Kusumoto (ed.), *Proceedings of the 26th North East Linguistics Society Conference (NELS)*, 87–100. Amherst, MA: University of Massachusetts.
- Han, Chung-hye & Dennis R. Storoshenko. 2012. Semantic binding of long distance anaphor *caki* in Korean. *Language* 88(4). 764–790. <https://doi.org/10.1353/lan.2012.0075>
- Han, Chung-hye, Dennis R. Storoshenko, Betty Leung & Kyeong-min Kim. 2015. The time course of long-distance anaphor processing in Korean. *Korean Linguistics* 17(1). 1–32. <https://doi.org/10.1075/kl.17.1.01han>
- Kaiser, Elsi. 2003. The quest for a referent: A crosslinguistic look at reference resolution. Philadelphia, PA: University of Pennsylvania dissertation.
- Kaiser, Elsi, Jeffrey T. Runner, Rachel S. Sussman & Michael K. Tanenhaus. 2009. Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition* 112(1). 55–80. <https://doi.org/10.1016/j.cognition.2009.03.010>
- Kuno, Susumu. 1987. *Functional syntax: Anaphora, discourse and empathy*. Chicago: University of Chicago Press.
- Lees, Robert B. & Edward S. Klima. 1963. Rules for English pronominalization. *Language* 39(1), 17–28. <https://doi.org/10.2307/410759>

- Moynes, John Abel. 1971. Reflexive and emphatic. *Language* 47. 141–163.  
<https://doi.org/10.2307/412192>
- Reinhart, Tanya & Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24 (4). 657–720.
- Sohng, Hong-Ki. 2004. A minimalist analysis of X<sup>0</sup> reflexivization in Chinese and Korean. *Studies in Generative Grammar* 14 (3). 375–396.
- Tenny, Carol. 2003. *Short distance pronouns, argument structure, and the grammar of sentience*. Manuscript, Carnegie Mellon University.
- Yoon, Jeong-Me. 1989. Long-distance anaphors in Korean and their cross-linguistic implications. In Caroline Wiltshire, Randolph Graczyk & Music Bradley (eds.), *Proceedings of the 25th Annual Meeting of the Chicago Linguistic Society*, 479–495. Chicago: Chicago Linguistic Society.



# A multi-dimensional approach to classification of Iran's languages

Erik Anonby<sup>1,2</sup>, Amos Hayes<sup>1</sup> and Robert Oikle<sup>1</sup>

<sup>1</sup>Carleton University / <sup>2</sup>Universität Bamberg

The enterprise of classification is central to the construction of a language atlas, particularly in the Iranian context. While existing two-dimensional models of language classification are useful as a starting point, they are ultimately incapable of handling some of the important complexities found in Iran's language situation. To address these issues, we propose a multi-dimensional "language relation web", based on a force-directed multigraph visualization, as an alternative model for expressing connections between language varieties. This architecture allows for differentiation and representation of multiple types of linkages, each of which constitutes a dimension of classification, in a single visualization: shared genealogical inheritance, structural similarity through contact, and association through ethnic identification. The resulting model provides new insights into the classification of Iran's languages and raises questions and prospects for the broader classification process.

**Keywords:** language classification, language atlas, languages of Iran, Iranian (Iranian) languages, tree model, wave model, 3D models, multi-dimensional models, directed graph, force-directed multigraph visualization.

## 1. Introduction

Language atlases typically approach language mapping from two directions: distribution of language varieties – language families, languages, dialects – in geographic and social space; and distribution of linguistic forms associated with these varieties: lexicon, phonology and grammar. Both approaches are important, and each of them refines our understanding of the other.

But which language varieties should be selected for an atlas? How should they be identified, represented and grouped? Such questions are fundamental to any language map. However, in the case of Iran's languages, attention to these questions is essential in overcoming deeply-rooted obstacles to the production of an atlas for the country's languages.

## 1.1 Language mapping and atlases of Iran

In a recent paper, we gave a historical overview of important efforts, starting in the 1950s, to build an atlas of Iran's languages. Unfortunately, none of these projects has yet resulted in the publication of an atlas. Several individual language maps make important contributions (*Atlas Narodov Mira* 1964; *TAVO* 1988; *Compendium* 1989; Izady 2006–2013; Windfuhr 2009; *Irancarto* 2011), but also face issues of reliability and detail (for discussion and critique, see Anonby 2015 and Anonby, Taheri-Ardali & Hayes 2019).

How is it that, after more than a century of linguistic research in Iran, there is still no satisfactory map, let alone atlas, of the country's languages? The same (Anonby 2015) paper points to a number of reasons: the complexity of the language situation, with hundreds of distinct varieties from several phyla; fragmentary documentation; issues of logistics and project design; limited dissemination of project results; and limited cooperation among scholars working toward this common goal. However, the study also shows that disagreement on the identification and classification of Iran's languages – much of it based on extralinguistic rather than linguistic factors – has been a major, ongoing obstacle (see also Anonby, Sabethemmatbadi & Hayes 2016).

## 1.2 The *Atlas of the Languages of Iran* (ALI)

After five years of planning, work on the *Atlas of the Languages of Iran* (ALI) (test version: <http://iranatlas.net>; Anonby, Taheri-Ardali et al. 2015–2020) was initiated in 2014 in response to these challenges. The overarching purpose of this research programme is to enable work toward a systematic understanding of the language situation in Iran. This is achieved through exploration of four key themes (Anonby, Taheri-Ardali & Hayes 2019; Taheri-Ardali et al. 2019):

- Linguistic and areal typology: What are important linguistic features of Iran's languages and dialects, and how are they distributed geographically?
- Language distribution: Where are these language varieties spoken, and how does this compare to the distribution of linguistic features?
- Language classification: How do scholars and speakers classify these language varieties, and how can scholarly classifications be improved?
- Language documentation: A record of the linguistic situation in Iran in the face of a rapid decline in linguistic diversity.

In the ALI research programme, an atlas of the country's languages is being developed by an international team of over 60 volunteer scholars and students (<http://iranatlas.net/module/atlasteam>). This atlas, which includes each of Iran's some

60,000 cities and villages, brings together existing publications and new data. It is capable of remote contributions by scholars and popular users and moderation of input by atlas editors. Because ALI brings together the work of many different people, it provides references to each data source, whether published work, collaborator field notes or user contributions. Fundamental to the purpose of the Atlas, it is designed to facilitate comparison of language distribution maps with maps based on attested linguistic forms (Anonby 2017; Anonby, Taheri-Ardali & Hayes 2019). ALI is being designed and built by the authors of this article along with Jean-Pierre Fiset at GCRC (Geomatics and Cartographic Research Centre, Carleton University: <https://gcr.ccarleton.ca>) using the open-source Nunaliit Atlas Framework (<http://nunaliit.org>). Nunaliit, which comes with a ready-made atlas template (GCRC 2013, Hayes et al. 2014, Hayes & Taylor 2019), is an innovative platform that has enabled the development of new approaches to language mapping (Anonby, Murasugi & Domínguez 2018). The language mapping functionalities developed in the present research program, presented in Anonby & Hayes (2016), are continuously incorporated into the Nunaliit platform and made freely available to other scholars through the website and on GitHub.<sup>1</sup>

Because of the complexity of the language situation in Iran – and the attendant issues (mentioned above) of incomplete knowledge about what language varieties exist; fragmentary documentation of these varieties; competing classifications; and disagreements about the status of specific varieties as ‘languages’ worth listing and documenting – we have designated language classification as one of the major themes in the Atlas, and it is central to the design and implementation of the research.

### 1.3 The structure of this chapter

This article investigates the methodology, role, and nature of the classification process for the *Atlas of the Languages of Iran* (ALI) research programme. After introducing key questions and issues in the research process (§ 1), we describe the work that has gone on in the construction of classification trees for the languages of Iran (§ 2). We then review other models of classification and consider their relative merits for treatment of the languages of Iran (§ 3). Observing that some of the limitations for the application of existing models for the classification of Iran's languages are associated with their two-dimensional (2D) conceptualizations, we submit a

---

1. The common Nunaliit and D3 integration code, as well as semantic relationships between atlas documents, are available at <https://github.com/GCRC/nunaliit>. Custom code related specifically to each ALI module (page) can be accessed by viewing the `nunaliit_custom.js` file with browser developer tools.

multi-dimensional classification model, which we label a ‘language relation web’, as an alternative way of expressing connections between language varieties (§ 4). We explore the consequences of a multi-dimensional representation for the languages of Iran and illustrate its functioning through differentiation and description of three prevalent ‘dimensions’ – types of linkages between languages: relation through genealogical<sup>2</sup> inheritance; structural similarity through language contact; and association through ethnic identification (§ 5). In the conclusion, we reflect on the contribution of the language relation web, and multi-dimensional representations generally, to an understanding of the nature of the broader classificatory process; and on prospects for the application and refinement of this model (§ 6).

## 2. An initial working classification of Iran’s languages

In order to address each of the issues identified in the introduction (§ 1), from an early stage in the Atlas project we have laid the groundwork for classification of Iran’s languages through the inventory of language varieties and construction of initial classification trees for each language family found in the country. We have reflected on the problematic nature of identification of certain varieties as ‘languages’, and we have considered the contributions as well as the limitations of this initial classification. Each of these elements is explored here.

### 2.1 Inventory of language varieties

Our first step toward a comprehensive classification of Iran’s languages was the establishment of a constantly expanding inventory of all language varieties in Iran at all levels – language families, languages, dialects. The inventory includes all labels we have found in published sources as well as varieties that the Atlas team has encountered so far through fieldwork (1. Hormozgan: Mohebbi Bahmani et al. 2015; Anonby 2016; Anonby & Mohebbi Bahmani 2016; 2. Kordestan: Mohammadirad et al. 2016; Anonby et al. 2019; 3. Chahar Mahal va Bakhtiari: Taheri-Ardali et al. 2016; 4. Ilam: Gheitasi et al. 2017; Aliakbari et al. 2014; 5. Bushehr: Nemati et al. 2017; all of which are collected in Anonby, Taheri-Ardali et al. 2015–2020). Even at this early stage in the project, we have compiled a list of over 500 distinct varieties, of which about 400 belong to the Iranian family.

---

2. Following Haspelmath (2004: 222), we use the term ‘genealogical’ in place of the traditional term ‘genetic’ to avoid confusion with biological genetic relationships.

## 2.2 Construction of classification trees

In order to make sense of these labels, we constructed an initial working classification of all the varieties using a traditional 2-dimensional (henceforth 2D)<sup>3</sup> tree structure (available at <http://iranatlas.net/module/classification>) – an imperfect, but commonly applied model of language classification (see § 3 for further discussion).

We started assembling and organizing each language variety in Iran into one of several major language family trees: Indo-European (Iranic [= Iranian], Indic [= Indo-Aryan], Armenian); Turkic; Semitic; Kartvelian; Dravidian; and Sign Languages. For Iranic, we extended the tree beyond the borders of Iran to include all members of the family, since most Iranic varieties are found there in any case, and this detailed classification work provides an opportunity to reassess the family as a whole. Although the Atlas concentrates on languages of the present day, and little is known about the history of most of these languages, our work on Iranic also features trees for languages known from the Old Iranic and Middle Iranic historical periods. We built all of the trees by consulting foundational and/or complete published classifications (especially Stilo 1981, 2007; Windfuhr 1989, 2009; Bulut 2014; *Ethnologue* 2017; *Glottolog* 2017), as well as many other important articles in the literature for further commentary. Where scholars agreed, we used this as a basis for the initial working classification. In the (not infrequent) cases of disagreement, we took notes about how sources disagreed, and why. In contrast to some of the existing classifications, which provide only partial and general references to sources, we furnished a comprehensive list of sources for classification, as well as sources for the scholarly debate over elements of the classification using linked footnotes throughout the classification.

After consulting the literature, we incorporated all varieties encountered during Atlas-related fieldwork, based on observations and initial assessments made by field researchers and other members of the research team (see ‘Inventory of language varieties’ above).

Finally, we worked with a number of senior colleagues to review specific sections of the classification, as follows:

- Iranic: Joan Baart, Habib Borjian, Geoffrey Haig, Michael Mehrdad Izady, Carina Jahani, Thomas Jügel, Maryam Nourzaei, Hassan Mohebbi Bahmani, Jaffer Sheyholislami, Don Stilo, Gernot Windfuhr, and two anonymous Kurdish-speaking referees

---

3. Although tree diagrams are rendered in two dimensions, John Nerbonne (p.c. 2018) notes that in general, along with the force-directed graphs described in §§ 4–5 below, they do not neatly exemplify a ‘two-dimensional’ or ‘three-dimensional’ graph, because the edges denote relations, not properties. See fn. 6 for further discussion.



- Armenian: Don Stilo
- Turkic: Christiane Bulut
- Semitic: Sami Aydin, Ulrich Seeger, Gernot Windfuhr
- Kartvelian: Babak Rezvani, Don Stilo

### 2.3 Backgrounding assessments of ‘language’ vs ‘dialect’

Many classifications, as well as other types of linguistic discourse, make a basic distinction between ‘languages’ and ‘dialects’. Linguists sometimes seek to draw the line between the two language variety types based on structural similarity (for example, percentage of cognate items on a given wordlist) or functional criteria such as measures of inherent mutual intelligibility. A technical definition of ‘dialect’ is a sub-type of another language variety, but in practice, people – including most linguists – tend to define the difference between languages and dialects based on extralinguistic factors such as ethnic or social identity; the prestige of the language variety; the existence of a standardized script; recognition as an official language within a political entity; the size of the community of speakers; or the urban vs rural provenance of the language community (Windfuhr 1995; Moreau 1997: 120–124). Within Southwestern Iranian, Persian is a sister – not parent – to varieties such as Lori, Larestani, Bandari, Kumzari, Bashkardi, and Sistani, but all of these varieties are commonly referred to as ‘Persian dialects’ in popular discourse, and many linguists follow suit.

In the context of Iran, differences in the factors that people rely on lead to greatly differing lists of languages for the country. Whereas official administrative materials usually specify between four and seven languages (e.g., SJS 1986), scholars often cite dozens (TAVO 1988), and one source – *Ethnologue* (2017) – claims 80 distinct languages. Counts given by speakers of the country’s many languages fall between the two extremes, but mother-tongue speakers of Tehrani-type Persian consistently provide smaller lists than speakers of minority language varieties (Anonby, Sabethematabadi & Hayes 2016).

In short, there is no single, definitive list of Iran’s languages that will adequately address the diverse audiences of the Atlas, even though all of them might expect one. But assessments of ‘language’ vs ‘dialect’ are not central to the main purposes of the Atlas (§ 1), and because of this we have chosen instead to focus on assembling an ever-expanding list of language varieties, and exploring the ways that these varieties fit together into a single coherent picture (Anonby, Sabethematabadi & Hayes 2016; Anonby & Sabethematabadi 2019).

## 2.4 Contribution and limitations of this initial classification

Although significant effort has been invested in this initial 2D tree-based classification, and it is already contributing toward a unified picture of the language situation in Iran, it is only a starting point for the ALI classification process: it is not intended as a definitive or complete account, but rather as an accessible, comprehensive and constantly improving working classification, and a way for scholars in the field of Iranian linguistics to see the approximate position of any language variety in relation to others.

But is it enough? At the same time as this initial classification clarifies many of the issues associated with the language situation as a whole, it highlights areas of disagreement, and points to limitations in the representation that expresses the classification.

In order to begin to address such issues, in the following sections we review existing models of language classification, which are expressed in two dimensions (§ 3), and then explore ways in which a multi-dimensional model could address persistent issues and enable new possibilities for language classification (§§ 4–5).

## 3. Overview of existing language classification models

There are two major models of language classification, and both of them have been expressed using two-dimensional representations. The first major model – the ‘tree’ model – dates back to Schleicher’s work in the mid-19th century (Schleicher 1853). In their ideal form, classification trees depict historical differentiation of language varieties as understood through the application of the comparative method – demonstration of shared innovations (Leskien 1876) – to language data. This approach is described and criticized more fully in Aikhenvald & Dixon (2001), Heggarty (2014) and François (2014). In the latter study, it is schematized as in Figure 1 (where capital letters refer to language varieties and ‘p-’ represents a proto-language has given rise to these varieties):

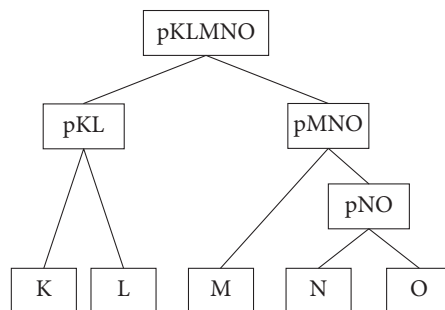
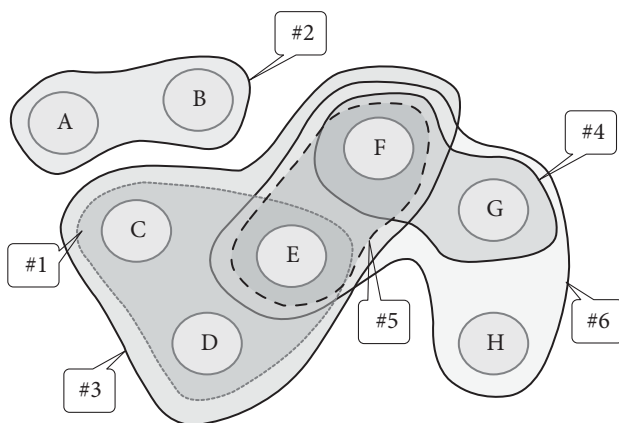


Figure 1. The ‘tree’ model of language classification (François 2014: 164)

As mentioned in § 2 above, this is the model that has been employed in ALI's initial working classification.

Already in 1872, Schuchardt and Schmidt had identified serious weaknesses in the tree model, and proposed a 'wave' model of classification as an alternative (Schmidt 1872). Like the tree model, the wave model in its ideal form is grounded in the comparative method. François (2014) provides a strong defense of the wave model's advantages over the tree model, particularly in terms of how the two models deal with diversification within languages, and intersecting subgroups – the fact that shared innovations often take place across genealogical lineages as a result of contact between varieties. In François' basic conceptualization of the wave model (Figure 2), capital letters refer to language varieties, and numbers refer to isoglosses that have resulted from shared innovations during the differentiation of these varieties.

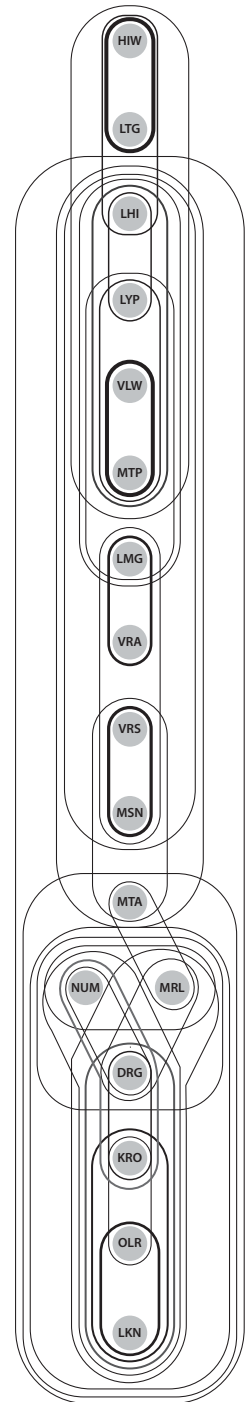
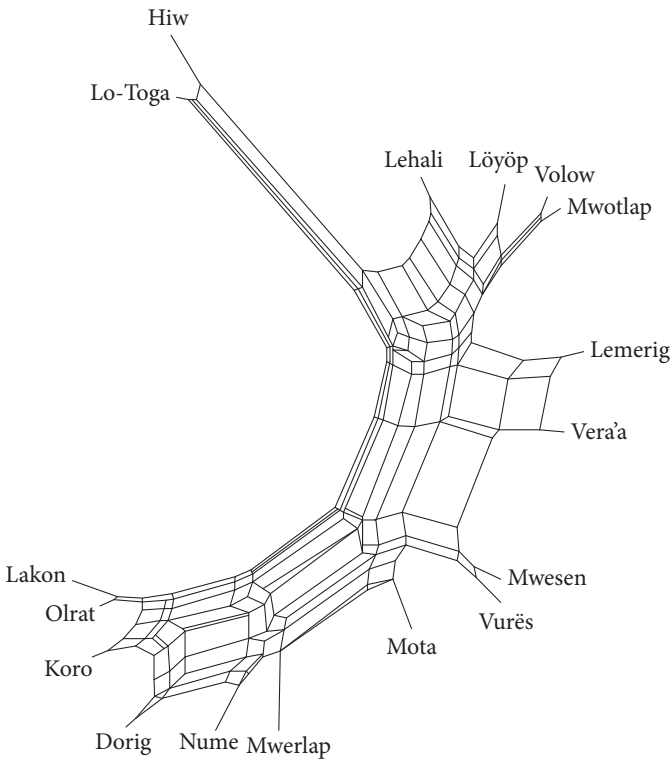


**Figure 2.** A conceptualization of the 'wave' model of language classification (François 2014: 169)

François further shows, and critiques, a 'NeighborNet' (pairwise distances) representation of the wave model (Figure 3) as it applies to a specific language situation in Vanuatu, and then proposes his own 'glottometric diagram' of the same situation (Figure 4). A detailed explanation of each of these approaches, and the resulting diagrams, is found in François (2014, 2017) and Kalyan & François (2018).

Of the various types of wave diagram, the glottometric diagram is the most sophisticated, and overcomes some of the weaknesses of the NeighborNet representation (see Gray et al. 2010; Heggarty et al. 2010), since it has the advantage of making explicit which specific differences and subgroupings are the result of shared innovations, rather than reflecting shared structures generally (François 2014: 179–180).<sup>4</sup>

4. However, as observed by John Nerbonne (p.c. 2018), the NeighborNet representation has the design-related and practical advantage of being automatically generated, whereas the hand drawn glottometric diagram is more difficult to replicate.



**Figures 3 and 4.** NeighborNet' (left) and 'glottometric' (right) visualizations of the 'wave' model of language classification (François 2014: 179, 183)

Finally, Jügel (2014) combines the tree and wave models in an innovative way in relation to West Iranic generally, and at the same time gives attention to effects of language contact (Figure 5; see also § 5 below).

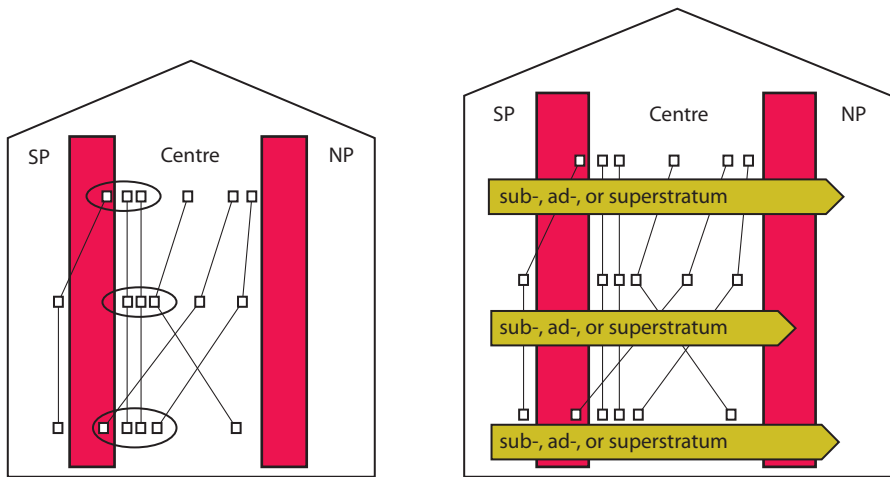


Figure 5. Combined tree/wave representations of West Iranic (Jügel 2014: 128)

The strengths and weaknesses of both tree and wave models have been debated for more than a century, and we will not attempt to resolve the discussion here, although we agree with François (2014) that the wave model is more faithful to realities of language differentiation and the role of language contact in language change. Still, despite the explanatory power of the wave model, the tree model still dominates scientific practice in comparative and historical linguistics: perhaps because of its visual simplicity; its convenience as an indexing tool; and the prevailing idea that language varieties are discrete units that can be identified and organized in relation to one another in a hierarchical taxonomy.

#### 4. A multi-dimensional language relation web for the languages of Iran

What would a detailed picture of the Iranian language family, with its hundreds of varieties, look like using any one of these representations?

Blažek & Schwarz (2017: 211ff.) bring together and schematize a heterogeneous assortment of tree-based lists and representations developed for Iranian (Windfuhr 2009; *Ethnologue* 2017; Gippert 2017; Jaxontov 2006; Cathcart 2015; and others), and provide further fine-grained assessments of their own (see especially the summary diagram on page 236). Araz (2017) draws a tree diagram for

Iranic that incorporates the populations of language groups. Other tree-based lists are *Glottolog* (2017), and the initial ALI classification introduced in § 2 above.

Detailed wave-model diagrams have never been constructed for the Iranic language family; however, Jügel's (2014) model addresses the historical development of six branches of West Iranic in a promising way. A detailed wave-based representation of this language family would also be feasible using the existing capabilities of NeighborNet, despite the limitations of this model.<sup>5</sup>The glottometric diagram of 17 closely related varieties in a continuum-like situation in Vanuatu is well-grounded, but visually intricate. Could any of these diagrams handle several hundred varieties in an accurate and cogent way, and if so, what impressions and insights would they communicate?

In our work on classification inside ALI over the past five years, we have struggled to integrate the complexities of the language situation in Iran into any of the existing models. On the one hand, the need for a straightforward data indexing tool, as exemplified by the tree model, is clear. On the other hand, however, and in line with the wave-based models, the ALI classification needs to account for the prevalence of intersecting genealogical subgroups, and the effects of language contact more generally. It should give a holistic, easily interpretable visual overview of the language situation, something that the NeighborNet model (despite its weaknesses) achieves most easily when a high number of language varieties are considered.

Finally, beyond the capabilities of any of the existing models, the ALI classification framework should provide clarity about the nature of existing classifications of Iranic and other language families in Iran – many of which are not based on rigorous application of the comparative method; consider the relevance and importance of other prevalent ways of linking languages, some of which are very material in the Iranian context; and make these other types of links explicit.

Whether taken as an integrated group or, in some cases, even individually, these explanatory requirements necessitate a flexible geometry which cannot be adequately handled using any of the existing 2D models. We also sense that even the most intricately conceptualized of the models, namely the glottometric diagram, is itself not optimally served by a 2D representation. Jügel's (2014) combined tree/wave model, with its crossing lines and intersecting contact language strata, is already, in fact, pushing toward a three-dimensional (3D) representation.

To tackle these issues, we have worked since 2015 as a cross-disciplinary team of linguists, designers and technical staff at GCRC (see § 1) to explore possibilities

---

5. To be precise, NeighborNet representations of the Indo-European phylum as a whole have been generated – and criticized (see Gray et al. 2010) – but members of the Iranic family are poorly represented there.

for conceptualization and visualization of language relation networks. After reviewing the open-source library of network models available through D3 (Data Driven Documents: <https://d3js.org>; Bostock 2019), we selected a directed graph, with a force-directed visualization (<http://bl.ocks.org/mbostock/1153292>), as the basis for classificatory model. A force-directed graph operates by linking a specified configuration of graph nodes to one another, and then, governed by this geometric constraint, by pushing each of the graph nodes away from all others, and from the frame, with a specified force (Grandjean 2015). For force-directed graphs with many nodes and a spherical frame, the resulting diagram is a spherical object.

We then customized this force-directed graph model to handle language classification. Language varieties serve as the graph nodes, and each node displays all the Atlas information relevant to that variety: all attested Persian and roman script labels, local pronunciation of the language variety name, where it is spoken, bibliographic references, etc. We have specified the direction of language relationships (general category vs category member) and made explicit multiple dimensions – types of relationships between language varieties (this latter enhancement is described in § 5 below). The specification of several types of links between nodes is known as a ‘multigraph’ feature (Gross & Yellen 2003).

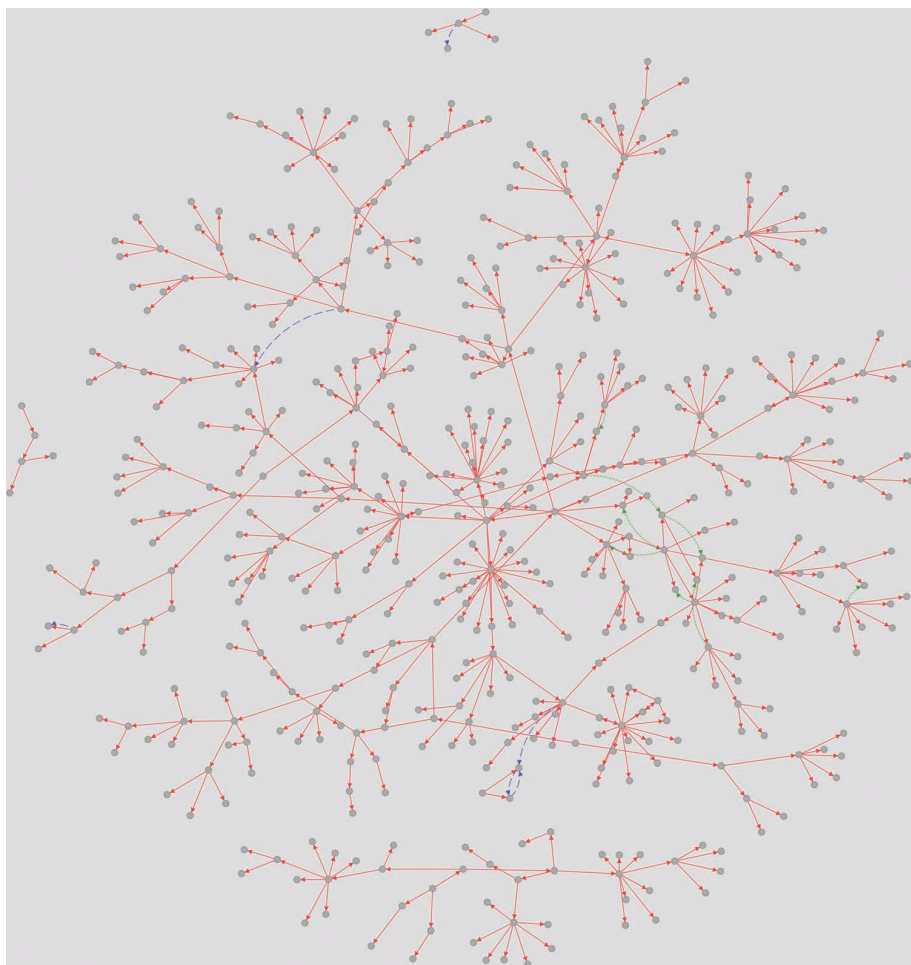
We have named the resulting representation a ‘multi-dimensional language relation web’.<sup>6</sup> When the initial ALI 2D classification tree (§ 2) content is migrated into a force-directed graph representation, here is the multi-dimensional relational structure that emerges (Figure 6):

This visualization provides a coherent and unified overview of the languages of Iran, and how they fit together, for the first time.

The graph is flexible, and can be turned and manipulated on the web page where it is displayed, so that the 3D architecture of the representation is evident. Other than the basic links that the underlying structure of the graph provides, the language varieties do not have a fixed placement within the graph. If any dot is moved to another location, the entire graph rebalances itself.

---

6. In earlier presentations of our work, including Anonby’s paper at NACIL1 (see Acknowledgements below) and Anonby (2017), we referred to this model as a ‘three-dimensional (3D) language relation web’. The force-directed graph produces a 3D visualization but, as John Nerbonne (p.c. 2018) has pointed out, the data themselves do not express fixed Cartesian coordinates. Instead, as with most graphic representations of language classification, the key features of the model are properties of ‘edges’ – the links between nodes – rather than properties of the nodes themselves. Along with the labels that accompany the nodes already, a more strictly defined third dimension could still be incorporated, with quantified structural properties replacing the force graph as an organizing principle. The advancement of time could also be incorporated into the visualization.



**Figure 6.** The languages of Iran as represented in a multi-dimensional language relation web (under development, from <http://iranatlas.net/module/taxonomy.selectMap>)

When labels are shown along with the language variety nodes (Figure 7), which can be done by clicking a ‘Node labels’ option above the graph, the representation is ostensibly more informative, but the density of the text obscures the underlying structure of the graph – a fitting metaphor for the actual complexity of the language situation!

A selection of language varieties from this web (for example, a particular language family) is more useful, so that the text does not obscure the relations between the varieties. A number of selections from the web are shown in the remainder of the chapter.



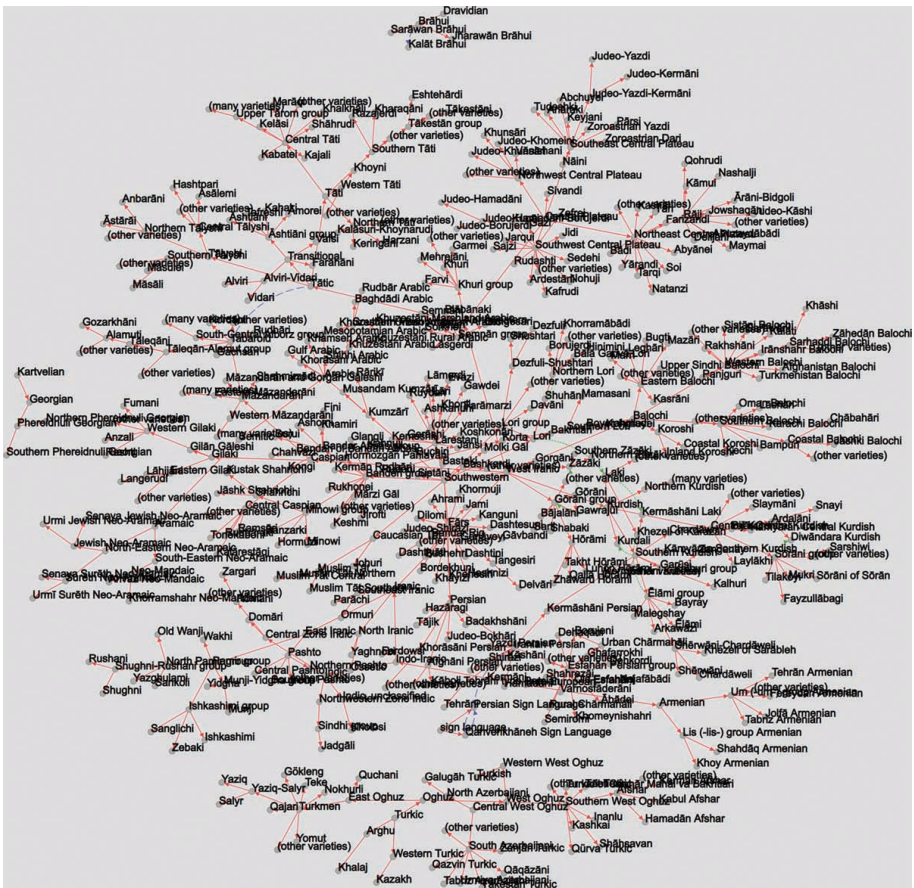


Figure 7. The languages of Iran in a multi-dimensional language relation web with labels (under development, from <http://iranatlas.net/module/taxonomy.selectMap>)

Further characteristics and enhancements of our language relation web for Iran, especially in regard to various dimensions – types of links between language varieties – are explored in the following section (§ 5). We then reflect on the contribution of the language relation web to the ALI research programme, and explore further enhancements and insights that could result from its application to language classification (§ 6).

## 5. Differentiation and visualization of links between language varieties

Along with the shift to a dynamic 3D graphic representation – but dependent on it – a basic enhancement in the language relation web proposed here is the explicit differentiation, and visualization, of several types of links between language varieties, which serve as additional dimensions. Three key link types that we use in the Atlas are:

- linguistic relation through genealogical inheritance;
- structural similarity through language contact; and
- association through ethnic identification.

In the language relation web, it is possible to show these types of links individually, or together as part of a multi-faceted picture of the language situation.

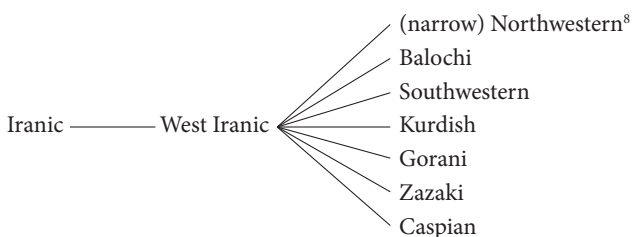
### 5.1 Linguistic relation through genealogical inheritance

In keeping with best practices for both tree and wave models, we maintain genealogical inheritance, established through the comparative method, as a central means of organizing language varieties in the Atlas classification. However, migration of the content to a dynamic 3D visualization avoids some of the drawbacks associated with the representation of genealogical relation in two dimensions.

To give an example from the Iranian family, a possible 2D tree classification of West Iranian, expressing one classificatory viewpoint in the current literature, would be as follows (Figure 8):<sup>7</sup>

---

7. To provide some background to this classification, West Iranian has traditionally been divided into Southwestern (Persian and close relatives) and Northwestern (everything else) (Tedesco 1921; followed by, e.g., Oranskij 1979; Schmitt 1989, 2000; Windfuhr 1989, 1995, 2009). However, there is increasing recognition that the many subfamilies of this wider Northwestern group have not been demonstrated to be more closely related to each other than to Southwestern (MacKenzie 1961; Paul 1998; Korn 2003, 2016; Borjani 2015, 2018). Jügel (2014; see also Figure 5) explores the idea of a heterogeneous 'Central' branch between Northwestern and Southwestern. In the broadest form of the current view, there are therefore several major branchings within West Iranian. In any case, the observations made about the static 2D vs dynamic multi-dimensional representations do not depend on the exact classification of West Iranian.



**Figure 8.** A 2D tree representation of West Iranian

The static and linear nature of this representation implies (intentionally or not)<sup>9</sup> a fixed, consecutive ordering of the branches: for example, that (narrow) Northwestern and Caspian are more distant from each other than Northwestern and Balochi. Yet West Iranian is more like a tangled continuum (Paul 1998; Stilo p.c. 2017) or circle (MacKenzie 1961: 75; Borjani p.c. 2014, p.c. 2017), with many criss-crossing links; and there has been no clear demonstration, as of yet, about relations between the highest-level branches.

In a dynamic 3D representation of West Iranian, generated using a force-directed graph (see § 4 above), the branches are spread evenly around the West Iranian node in all directions, with no intrinsic ordering intended or implied (Figure 9). If one of the branches is moved to another location by the viewer, the graph automatically rebalances in space. A satisfactory replication of the dynamic, 3D characteristics of this digital representation is not possible here, but it is available through the ALI website.<sup>10</sup> A screenshot of this selected portion of the language relation web shows one possible configuration:

---

8. Building on the preceding note, and inspired by Stilo (p.c. 2014, p.c. 2017), the use of the label ‘Northwestern’ here refers to a narrow (Stilo uses the term ‘core’) Northwestern group which includes Tatic, Central Plateau, and Semnani, and treats Balochi and Kurdish as distinct high-level branches within Iranian. Stilo also treats Gorani, Zazaki and Caspian as part of ‘core Northwestern’; Borjani (2009, 2011) posits a more restricted ‘Median’ group that brings together only Tatic and Central Plateau.

9. According to the principles of diachronic linguistics, the visual ordering of the daughters is not inherently meaningful, but in practice it often is intended as such, and static representations themselves encourage inferences of intentional ordering.

10. To access and interact with this representation, go to the language relation web page (<http://iranatlas.net/module/taxonomy.selectMap>) and select the ‘West Iranian (sample)’ in the white dropdown menu in the lower middle section of the navigation bar. To see the labels of each of the language variety nodes, check the ‘Node labels’ box.

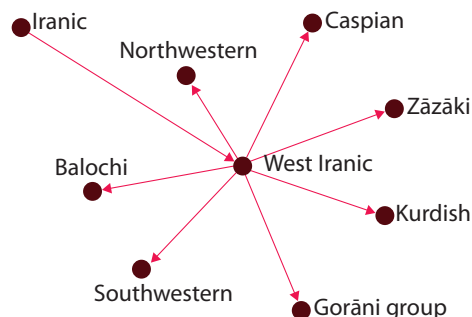


Figure 9. Representation of West Iranic in the language relation web

It would be equally fruitful to produce such a representation of Indo-European as a whole, where the parent – Proto-Indo-European – is very clear, but relations between the major branches are ambivalent (Ringe et al. 2002: 98; see also the diagrams in Gray et al. 2010: 3927 and Bouckaert et al. 2013).

Visualization of mixed languages is another important contribution of the language relation web to the representation of genealogical relations. Here, we are referring to mixed languages in the technical, narrow sense of the term, that is, languages constituted by inheritance of foundational, equivalent components from more than one genealogical parent (cf. Matras & Bakker 2003). In relation to Iranic, Kumzari has been described as a mixed language with Southwestern Iranic and Arabic components (van der Wal Anonby 2015). In the language relation web, a mixed language is shown with genealogical links to both of its parents (Figure 10); in doing so, it brings together portions of the classification from two different language varieties – and in the case of Kumzari, two unrelated language families.

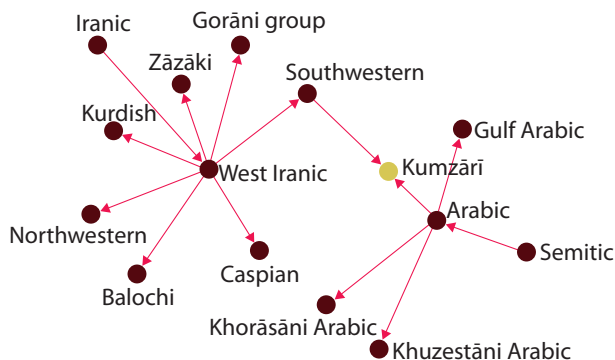


Figure 10. Language relation web selection, Kumzari: Representation of a mixed language

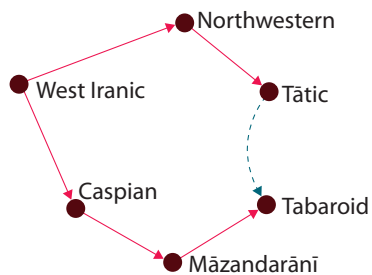
## 5.2 Structural similarity through language contact

In many cases, longstanding contact between languages – with or without close genealogical relation – results in structurally intermediate varieties. (This differs from the case of mixed languages discussed immediately above, where a language has foundational, equivalent components inherited from more than one parent language.)

In the Iranian context, this situation is exemplified by a group of language varieties in the western Alborz Mountains, referred to as Tabaroid (Borjjan 2013a, 2013b, p.c. 2014) or Tatoid (Stilo 2018), which are structurally transitional between Mazandarani (in the Caspian branch of West Iranic) and Tatic (part of the narrow Northwestern branch of West Iranic).

While a careful application of the comparative method might point to genealogical affiliation with Mazandarani, as Borjjan maintains (but cf. Stilo 2018), a classificatory representation that ignores the contribution of Tatic as a contact language (whether as a superstrate, adstrate, or substrate) does not adequately capture the structural similarity between Tabaroid/Tatoid and Tatic at all levels of the language.<sup>11</sup>

In the language relation web (Figure 11), the solid red lines indicate the genealogical relation of Tabaroid/Tatoid and its Caspian parent Mazandarani. In addition, the dashed blue line from Tatic to Tabaroid/Tatoid acknowledges the contribution of Tatic as a contact variety, and brings together Caspian and Tatic in this area of the classification.



**Figure 11.** Language relation web selection, Tabaroid: Genealogical relation  $\neq$  structural similarity through contact

11. Stilo (p.c. 2017) views Southern Taleshi (= Southern Talyshi), which is more easily intelligible to nearby forms of Tati than to Central or Northern Taleshi, as a better example of a structurally transitional variety. The same principles raised in the discussion of the importance of a ‘structural similarity through contact’ link for Tabaroid/Tatoid would apply to this situation.

This representation is simplistic, since neither the degree of similarity between varieties, nor the extent of contact influence, is a binary feature (present vs absent / linked vs unlinked). Certainly the degree of similarity between individual varieties – whatever the genealogical relation between them – can be quantified, and we plan to do this once our lexical, phonological and morphosyntactic datasets (recently featured and analyzed in Anonby & Taheri-Ardali 2017) become more complete.

A further shortcoming of this representation, in its present form, is that it does not specify the particular type of contact influence (superstrate/adstrate/substrate), each of which tends to pattern in a specific way (Thomason & Kaufman 1988). Jügel (2014) and Stilo (p.c. 2017) emphasize the fact that substrate phenomena, brought about by large-scale language shift, are prevalent in the context of Iran. These important distinctions, which would depend on a quantifiable typology of language contact, are beyond the scope of our current research.

Still, we submit this representation as an illustration of how recognition of structural similarity resulting from language contact can enhance existing methods of classification, and clarify our understanding of local and more general characteristics of the language situation.

### 5.3 Association through ethnic identification

A third kind of linkage between languages in the language relation web is that of association through social factors. For Iran in particular, we have observed the importance of ethnic identification as a way of organizing language varieties (Anonby & Sabethematabadi 2019). This type of link is not often acknowledged explicitly in the literature, but in the context of Iran it may be as important as genealogy and shared linguistic structures are, in the way that people – linguists as well as others – classify languages.

As discussed in § 2, linguists often refer (or at least claim to refer) to empirical measures such as bundling of shared historical innovations, lexical similarity, or degree of inherent intelligibility to group language varieties together and to label them as a ‘language’, ‘dialect’, or another similar term. Because of this appeal to the possibility of measurement, linguists working on the languages of Iran might intuitively view their assessments as factual and sufficient.

In actual fact, few conclusive results have emerged from the applications of the comparative method to the main divisions of West Iranian, and few studies acknowledge the massive variation within language varieties that people group together as unique ‘languages’ and ‘dialects’ (Matras et al. 2016 is an important exception). Perhaps most comparative work on Iranian has depended on lexicon,

but the pervasive and longstanding influence of Persian, Kurdish and other major regional languages, as well as widespread local-level contact phenomena, have led to massive borrowing of vocabulary, so that historical studies of lexicon cannot be conclusive on their own.<sup>12</sup> And almost no research that systematically considers the role of (either reported or measured) mutual intelligibility in grouping language varieties together has been conducted on the languages of Iran (for two modest exceptions, see Anonby (2003) and Anonby & Yousefian (2011)).

However, the literature is full of cases where linguists define and group language varieties based on intangible social factors as such as ethnic identification, and background the centrality of measures of historical, structural or functional similarity. To give a very basic example, most linguists refer to Semnani as a dialect of Persian, even though it is certainly not descended from Old or Middle Persian; and Semnani may in fact be best described as a *Sprachbund* – a bundle of historically distinct varieties that have developed shared structures – rather than as a single ‘language’. Another example, and one that is more frequently debated in the literature, is the fact that many linguists continue to refer to Zazaki and Gorani, linguistically, as types of Kurdish or members of a larger Kurdish family (Hassanpour 1992, 1998, 2012; Windfuhr 2009; Kurdish Academy of Language 2014, 2017). Yet many important studies have pointed out that the evidence for a close genealogical connection between Northern and Central Kurdish on the one hand, with Zazaki and Gorani on the other hand, is ambivalent at best (MacKenzie 1961; Paul 1998, 2008; Haig & Öpengin 2014; Jügel 2014).<sup>13</sup>

Is it important, or even relevant, to take socially-defined views of language relation into consideration? In fact, as pointed out in Hassanpour (1998), Anonby & Sabethematabadi (2019) and in § 2 above, language classification – the identification of language varieties, defining them as a certain type of variety (‘language’, ‘dialect’, etc.), and the organization of these varieties into groups – is a socially-driven process, and is intimately tied to language identity (whether in relation to one’s own languages, or the languages of others). Ultimately, the way in which speakers identify and classify their own language varieties can be a determining factor in the way that languages evolve.

Therefore the problem is not that linguists and others depend on social factors in their classifications, but that linguists acknowledge neither the importance of these social factors, nor their distinctness from genealogical relation and structural similarity, in their classificatory assessments. Again, as Thomas Jügel (p.c. 2016)

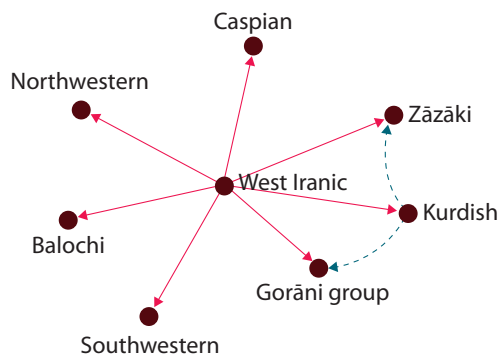
---

12. Heggarty, Anderson et al.’s (2015–2020) ongoing comparative lexical work on Indo-European, which has a strong representation of West Iranian varieties, does however attempt to control for effects of lexical borrowing in the reconstruction of language relations.

13. Likewise, within this classificatory debate, recurrent claims about a close relationship between Zazaki and Gorani have yet to be substantiated.

has succinctly stated: “Linguists have to understand that they cannot tell people who they are, and language communities should understand that [genealogical] language affiliation is not the same as identity affiliation”.

For the case of Kurdish, Zazaki and Gorani (Figure 12), the configuration of solid red lines in the language relation web indicates that a close genealogical relation has not been established for these West Iranian varieties. However, the additional dotted green lines, which indicate association based on ethnicity, link Zazaki and Gorani with Kurdish and pull these varieties together in the graph.



**Figure 12.** Language relation web selection, Zazaki / Gorani: Genealogical relation  $\neq$  ethnic identification

This particular representation of Zazaki, Gorani and Kurdish may not satisfy all audiences, but its meaning is clear, and linguists as well as speakers can begin to see how – and why – each language variety might be classified in a certain way by others.

## 6. Summary and prospects

This article explores the methodology, role, and nature of the classification process for the *Atlas of the Languages of Iran* (ALI) research programme. We show that the enterprise of classification is central to the construction of a language atlas, particularly in the context of Iran. We argue further that while two-dimensional (2D) models of language classification are useful as a starting point, they are ultimately incapable of handling many of the complexities found in Iran's language situation in a satisfactory way.

To address these issues, we propose a multi-dimensional language relation web, based on a force-directed multigraph visualization, as an alternative model for expressing connections between language varieties. This model results in a coherent visualization that is simultaneously detailed, accessible and interpretable. In our application of the model to the languages of Iran, we distinguish and describe



three dimensions – three types of linkages between varieties which are essential to make explicit for diverse audiences in any language atlas: relation through genealogical inheritance; structural similarity through language contact; and association through ethnic identification. In this way, disparate parts of the classification that are in some way connected are brought together in a multigraph, and the types of links between language varieties – both intuitive and pervasive but only erratically acknowledged in the literature – are made explicit.

The model we propose in this study is not intended as a definitive account of the workings of language relations, but as an inquiry into how models of classification could be improved, and nuanced, by rethinking their basic geometry. Certain dimensions, such as the three that we have identified here, are more central or valuable in understanding relations between language varieties, but in this multigraph-based model there is in fact no limit to the number of dimensions that could be expressed. Further, each of the three dimensions here could be subdivided: competing genealogical classifications could be differentiated and represented together in the same web; lexical, phonological and morphosyntactic similarity could be distinguished; and for ethnic identification, differences in outlook between ethnic groups, subgroups within a larger ethnicity, or even differences between individual perspectives, could (and should) be made explicit and distinguished.

Currently, ALI-related colleagues at the GCRC are working on two further classification-related Atlas design tasks. First, they are computationally integrating the language relation web with the language distribution maps and language data maps, so that when any modification is made to the language relation web, it will update the language maps automatically. Secondly, other ALI team members are investigating how different (but overlapping) groups of people (Iranian linguists; Western linguists; mother-tongue speakers of Persian; and speakers of specific minority languages in Iran), and individuals within these groups, view the language situation in Iran differently from one another (Anonby et al. 2016; Sabethematabadi et al. 2017; Anonby & Sabethematabadi 2019), and how their perspectives condition their expectations for language maps and correlate to the actual maps that they produce.

There are a number of other promising areas of further exploration. These include:

- a proper integration of the strengths of wave models, and particularly the combined tree/wave model (Jügel 2014) and the glottometric diagram (François 2014; Kalyan & François 2018), into multi-dimensional architecture;
- integration of quantified measures into the multi-dimensional modelling, for example, to express relative structural distance between language varieties using line length or thickness rather than only binary (present/absent) links;

- construction and comparison of separate webs for each type of linkage (genealogy; structural similarity; ethnicity) and, in the case of the first two types, further subdivision and comparison of the innovation- or structure-based webs generated for each linguistic domain (lexicon, phonology, morphosyntax);
- reflection on the place of functional measures such as inherent intelligibility, and social linkages beyond ethnicity, in a language relation web;
- experimentation with additional three-dimensional visualizations such as density cloud isosurfaces (Jardine 2017) to provide simple, powerful and representative graphic overviews of the language situation in Iran.

In conclusion, this exploration of the classification research process for the languages of Iran raises questions about the nature of classification generally, and unlocks new possibilities for refinement of classificatory models that better represent the complexities of real-world language situations.

## Funding

Various elements of the research were supported by SSHRC (Social Sciences and Humanities Research Council of Canada), Carleton University, Universität Bamberg and the Alexander von Humboldt Foundation.

## Acknowledgements

An earlier version of this chapter was presented by Erik Anonby at the 1st North America Conference on Iranian Linguistics (NACIL1), Stony Brook University, New York, April 28–30, 2017. Alongside the work of co-authors Hayes and Oikle, we wish to acknowledge the contribution of Jean-Pierre Fiset of GCRC (Geomatics and Cartographic Research Centre, Ottawa) in the design and technical implementation of the language relation web at the heart of this research. We are also grateful for detailed, incisive feedback provided by colleagues, in particular Chams Bernard, Habib Borjian, Alexandre François, Paul Heggarty, Thomas Jügel, John Nerbonne, Jaffer Sheyholislami, Parisa Sabethematabadi, Don Stilo and Mortaza Taheri-Ardali. All shortcomings in the present study are, of course, the responsibility of the authors.

## References

- Aikhenvald, Alexandra Y. & R. M. W. Dixon (eds.). 2001. *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press.
- ALI = Anonby, Erik, Mortaza Taheri-Ardali et al. 2015–2020 (see reference below).
- Aliakbari, Mohammad, Mojtaba Gheitasi & Erik Anonby. 2014. On language distribution in Ilam Province, Iran. *Iranian studies* 48(6). 1–16.

- Anonby, Erik. 2003. Update on Luri: How many languages? *Journal of the Royal Asiatic Society* 13: 2, 171–97.
- Anonby, Erik. 2015. Mapping Iran's languages: Situation and prospects. *6th International Conference of Iranian Linguistics (ICIL6)*, Ilia State University, Tbilisi, Georgia, June 23–26.
- Anonby, Erik. 2016. The Keshmi (Qeshmi) dialect of Hormozgan Province, Iran: A first account. *Studia Iranica* 44(2). 165–206.
- Anonby, Erik. 2017. Constructing a linguistic atlas for Iran's languages. *Bayerisches Orientkolloquium: Kulturlandschaften Südostanatoliens und Mesopotamiens: Lokale und transnationale Perspektiven* [Bavarian Orientalist Colloquium: Cultural Landscapes of South Anatolia and Mesopotamia: Local and Transnational Perspectives], Otto-Friedrich-Universität Bamberg, May 18.
- Anonby, Erik & Amos Hayes. 2016. Mapping endangered languages of Iran: Application of the Nunaliit atlas framework to language mapping. *International Symposium on Endangered Iranian Languages (ISEIL) 2016*, Institut National des Langues et Civilisations Orientales (INALCO) and Sorbonne Nouvelle Paris III, July 8–9.
- Anonby, Erik, Masoud Mohammadirad & Jaffer Sheyholislami. 2019. Kordestan Province in the *Atlas of the Languages of Iran*: Research process, language distribution, and language classification. In Songül Gündoğdu, Ergin Öpengin, Geoffrey Haig & Erik Anonby (eds.), *Advances in Kurdish linguistics*, 9–38. Bamberg, Germany: Otto-Friedrich-Universität Bamberg.
- Anonby, Erik & Hassan Mohebbi Bahmani. 2016. Shipwrecked and landlocked: Kholosi, an Indo-Aryan language in South-West Iran. *Studia Iranica* supplement, vol. 43, 1–23.
- Anonby, Erik, Kumiko Murasugi & Martín Domínguez. 2018. Mapping language and land with the Nunaliit atlas framework: Past, present and future. In Sebastian Drude, Nicholas Ostler & Marielle Moser (eds.), *Endangered languages and the land: Mapping landscapes of multilingualism. Proceedings of the 22nd Conference of the Foundation for Endangered Languages, Reykjavik, August 23–25, 2018*, 56–63. Online at: <http://www.e-publishing.org/docs/4/01/FEL-2018-09.pdf>
- Anonby, Erik & Parisa Sabethematabadi. 2019. Representing complementary user perspectives in a language atlas. In D. R. Fraser Taylor, Erik Anonby & Kumiko Murasugi (eds.), *Further developments in the theory and practice of cybercartography: International dimensions and language mapping*, 413–440. Amsterdam: Elsevier.  
<https://doi.org/10.1016/B978-0-444-64193-9.00023-3>
- Anonby, Erik, Parisa Sabethematabadi & Amos Hayes. 2016. *Reconciling contradictory perspectives of language identity in the 'Atlas of the Languages of Iran'* (Cognitive Science Colloquium Series), Ottawa: Carleton University, February 25.
- Anonby, Erik & Mortaza Taheri-Ardali. 2017. Linguistic diversity and language contact in Chahar Mahal va Bakhtiari Province, Iran. *7th International Conference on Iranian Linguistics (ICIL7)*, Lomonosov Moscow State University and the Institute of Linguistics of the Russian Academy of Sciences, Moscow, August 28–30.
- Anonby, Erik, Mortaza Taheri-Ardali et al. 2015–2020. *Atlas of the Languages of Iran (ALI)*. Ottawa: Geomatics and Cartographic Research Centre, Carleton University. Online at: <http://iranatlas.net>
- Anonby, Erik, Mortaza Taheri-Ardali & Amos Hayes. 2019. The 'Atlas of the Languages of Iran' (ALI): A research overview. *Iranian Studies* 52(1–2). 199–230. Online at: <https://doi.org/10.1080/00210862.2019.1573135>
- Anonby, Erik & Pakzad Yousefian. 2011. *Adaptive multilinguals: A study of language on Larak Island*. Uppsala: Acta Universitatis Upsaliensis.

- Atlas Narodov Mira 1964 = Bruk, S. I. & V. S. Apenchenko (eds.). 1964. *Atlas narodov mira* [Atlas of the World's Peoples]. Moscow: Glavnoye Upravleniye Geodezii i Kartografii.
- Araz (Wikipedia username). 2017. Iranian languages family tree. Online at: [https://en.wikipedia.org/wiki/Iranian\\_languages](https://en.wikipedia.org/wiki/Iranian_languages), [https://commons.wikimedia.org/wiki/File:Iranian\\_Family\\_Tree\\_v2.0.png](https://commons.wikimedia.org/wiki/File:Iranian_Family_Tree_v2.0.png), accessed 2 December, 2019.
- Blažek, Václav & Michal Schwarz. 2017. *Early Indo-Europeans in Central Asia and China: Cultural relations as reflected in language*. Innsbruck: Innsbrucker Beiträge zur Kulturwissenschaft.
- Borjjan, Habib. 2009. Median succumbs to Persian after three millennia of coexistence: Language shift in the Central Iranian Plateau. *Journal of Persianate Societies* 2.1. 62–87. <https://doi.org/10.1163/187471609X454671>
- Borjjan, Habib. 2011. Kashan, ix. The median dialects of Kashan. *Encyclopaedia Iranica* 16.1, 38–48. New York: Center for Iranian Studies, Columbia University. Online at: <http://www.iranicaonline.org/articles/kashan-ix-the-median-dialects-of-kashan>, accessed 2 December, 2019.
- Borjjan, Habib. 2013a. *Is there continuity between Persian and Caspian? Linguistic relationships in South-Central Alborz* (American Oriental Society Essay 13). New Haven: American Oriental Society.
- Borjjan, Habib. 2013b. Perso-Tabaric dialects in the language transition zone bordering Mazandaran. *Studia Iranica* 42 (2). 195–225.
- Borjjan, Habib. 2015. The Biabanak language group of the Central Desert of Iran. *International Symposium on Endangered Iranian Languages (ISEIL) 2015*, Goethe University, Frankfurt, February, 19–20.
- Borjjan, Habib. 2018. The dialect of Khur. In Matteo De Chiara, Adriano V. Rossi & Daniel Septfonds (eds.), *Mélanges d'ethnographie et de dialectologie irano-aryennes à la mémoire de Charles-Martin Kieffer* [Collected Contributions to Irano-Aryan Ethnography and Dialectology], 77–98. Cahiers de Studia Iranica 61. Paris: Association pour l'Avancement des Études Iraniennes.
- Bostock, Mike. 2019. D3 (Data driven documents). Online at: <https://d3js.org>
- Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill et al. 2013. Correction to: Mapping the origins and expansion of the Indo-European language family. *Science* 342. 6165, 1446.
- Bulut, Christiane. 2014. Turkic varieties in West Iran and Iraq: Representatives of a South Oghuz dialect group? In Heidi Stein (ed.), *Turkic in Iran: Past and Present*, 15–99. Wiesbaden: Harrassowitz.
- Cathcart, Chundra Aroor. 2015. Iranian dialectology and dialectometry. Berkeley, CA: University of California Berkley dissertation.
- Compendium = Schmitt, Rüdiger (ed.). 1989. *Compendium linguarum Iranicarum* [Compendium of Iranian languages]. Wiesbaden: Reichert.
- Ethnologue = Simons, Gary F. & Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the world*, 20th edition. Dallas: SIL International. Online at: <http://www.ethnologue.com>
- François, Alexandre. 2014. Trees, waves and linkages: Models of language diversification. In Claire Bowerman & Bethwyn Evans (eds.), *Routledge handbook of historical linguistics*, 161–189. London: Routledge.
- François, Alexandre. 2017. Méthode comparative et chaînages linguistiques: Pour un modèle diffusionniste en généalogie des langues [Comparative method and linguistic chaining: Toward a diffusionist model in language genealogy]. In Jean-Léo Léonard (ed.), *Diffusion: Implantation, affinités, convergence* [Diffusion: Implantation, affinities, convergence], 43–82. Mémoires de la Société de Linguistique de Paris, XXIV. Louvain: Peeters.

- GCRC (Geomatics and Cartographic Research Centre). 2013. *Nunaliit atlas framework*. Ottawa: GCRC, Carleton University. Online at: <http://nunaliit.org>
- Gheitasi, Mojtaba et al. 2016. Ilam Province. In: Anonby, Erik, Mortaza Taheri-Ardali et al. 2015–2020. *Atlas of the Languages of Iran (ALI)*. Ottawa: Geomatics and Cartographic Research Centre, Carleton University. Online at: <http://iranatlas.net/module/language-distribution.ilam>
- Gippert, Jost. 2017. Genealogical tree of Iranian languages. Online at: <http://titus.fkildg1.uni-frankfurt.de/didact/idg/iran/iranstam.htm>
- Glottolog = Hammarström, Harald, Sebastian Bank, Robert Forkel & Martin Haspelmath. 2017. *Glottolog 3.1*. Jena: Max Planck Institute for the Science of Human History. Online at: <http://glottolog.org>
- Grandjean, Martin. 2015. Introduction à la visualisation de données: l'analyse de réseau en histoire [Introduction to data visualization: Network analysis in history]. *Geschichte und Informatik* 18/19, 109–128.
- Gray, Russell D., David Bryant & Simon J. Greenhill. 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365. 3923–3933. <https://doi.org/10.1098/rstb.2010.0162>
- Gross, Jonathan & Jay Yellen. 2003. *Handbook of graph theory*. Boca Raton: CRC Press. <https://doi.org/10.1201/9780203490204>
- Haig, Geoffrey & Ergin Öpengin. 2014. Kurdish: A critical research overview. *Kurdish Studies* 2(2). 99–122. <https://doi.org/10.33182/ks.v2i2.397>
- Haspelmath, Martin. 2004. How hopeless is genealogical linguistics, and how advanced is areal linguistics? *Studies in Language* 28(1). 209–223. <https://doi.org/10.1075/sl.28.1.10has>
- Hassanpour, Amir. 1992. *Nationalism and Language in Kurdistan, 1918–1985*. San Francisco: Mellen Research University Press.
- Hassanpour, Amir. 1998. The identity of Hewrami speakers: Reflections on the theory and ideology of comparative philology. In Anwar Soltani (ed.), *Anthology of Gorani Kurdish poetry*, compiled by A. M. Mardoukhi (1739–1797), 35–49. London: Soane Trust for Kurdistan.
- Hassanpour, Amir. 2012. The indivisibility of the nation and its linguistic divisions. *International Journal of the Sociology of Language* 217. 49–73.
- Hayes, Amos, Peter L. Pulsifer & Jean-Pierre Fiset. 2014. The Nunaliit cybercartographic atlas framework. In D. R. Fraser Taylor (ed.), *Developments in the theory and practice of cybercartography: Applications and indigenous mapping*, 129–140. Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-444-62713-1.00009-X>
- Hayes, Amos & D. R. Fraser Taylor. 2019. Developments in the Nunaliit cybercartographic atlas framework. In D. R. Fraser Taylor, Erik Anonby & Kumiko Murasugi (eds.), *Further developments in the theory and practice of cybercartography: International dimensions and language mapping*, 205–218. Amsterdam: Elsevier.
- Heggarty, Paul. 2014. Prehistory through language and archaeology. In Claire Bowerman & Bethwyn Evans (eds.), *Routledge handbook of historical linguistics*, 598–626. London: Routledge.
- Heggarty, Paul, Cormac Anderson, et al. 2015–2020. *Indo-European: Comparison of relationships (IE-Cor)*. Jena: Max Planck Institute for the Science of Human History. Online at: <http://cobl.info>
- Heggarty, Paul, Warren Maguire & April McMahon. 2010. Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365. 3829–3843. <https://doi.org/10.1098/rstb.2010.0099>

- Irancarto 2012 = Hourcade, Bertrand et al. 2012. *Irancarto*. Paris and Tehran: CNRS / University of Tehran. Online at: <http://www.irancarto.cnrs.fr>
- Izady, Michael Mehrdad. 2006–2013. Linguistic composition of Iran. Online at: [http://gulf2000.columbia.edu/images/maps/Iran\\_Languages\\_lg.png](http://gulf2000.columbia.edu/images/maps/Iran_Languages_lg.png)
- Jardine, Kevin. 2017. Galaxy map. Online at: <http://galaxymap.org/drupal>
- Jaxontov, Sergei E. 2006. Leksikostatističeskaja klassifikacija iranskix jazykov [Lexicostatistical classification of the Iranian languages]. In Mikhail N. Bogoljubov (ed.), *Indo-iranskoe jazykoznanie i tipologija jazykovyx situacij* [Indo-Iranic linguistics and the typology of language situations], 93–101. St. Petersburg: Nauka.
- Jügel, Thomas. 2014. On the linguistic history of Kurdish. *Kurdish Studies* 2(2). 123–142. <https://doi.org/10.33182/ks.v2i2.398>
- Kalyan, Siva & Alexandre François. 2018. Freeing the comparative method from the tree model: A framework for historical glottometry. In Ritsuko Kikusawa & Lawrence Reid (eds.), *Let's talk about trees: Tackling problems in representing phylogenetic relationships among languages* (Senri Ethnological Studies 98), 59–89. Osaka: National Museum of Ethnology.
- Korn, Agnes. 2003. Balochi and the concept of North-west 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. 2016. A partial tree of Central Iranian. *Indogermanische Forschungen* 121. 401–434. <https://doi.org/10.1515/iff-2016-0021>
- Kurdish Academy of Language. 2014. KAL's Kurdish genealogy tree. Online at: <http://kurdish-academy.org/?q=node/159>, accessed 2 December, 2019.
- Kurdish Academy of Language. 2017. The dialects of Kurdish language. Online at: <http://kurdish-academy.org/?q=node/212>, accessed 2 December, 2019.
- Leskien, August. 1876. *Die Declination im Slawisch-Litauischen und Germanischen* [Declination in Slavic-Lithuanian and Germanic]. Leipzig: Hirzel.
- MacKenzie, David N. 1961. The origins of Kurdish. *Transactions of the Philological Society* 1961. 68–86. <https://doi.org/10.1111/j.1467-968X.1961.tb00987.x>
- Matras, Yaron & Peter Bakker (eds.). 2003. *The mixed language debate: Theoretical and empirical advances*. Berlin: de Gruyter. <https://doi.org/10.1515/9783110197242>
- Matras, Yaron et al. 2016. *The dialects of Kurdish*. Manchester: University of Manchester. Online at: <http://kurdish.humanities.manchester.ac.uk>
- Mohammadirad, Masoud et al. 2016. Kordestan Province. In: Anonby, Erik, Mortaza Taheri-Ardali et al. 2015–2020. *Atlas of the Languages of Iran (ALI)*. Ottawa: Geomatics and Cartographic Research Centre, Carleton University. Online at: <http://iranatlas.net/module/language-distribution.kordestan>
- Moreau, Marie-Louise. 1997. *Sociolinguistique: Concepts de base* [Sociolinguistics: Basic concepts]. Wavre: Mardaga.
- Oranskij, Iosif M. 1979. *Les langues iraniennes* [The Iranian languages]. Paris: Klincksieck.
- Paul, Ludwig. 1998. The position of Zazaki among West Iranian Languages. In Nicholas Sims-Williams (ed.), *Proceedings of the Third European Conference of Iranian Studies held in Cambridge on 11th to 15th September 1995* (Part I: Old and Middle Iranian Studies), 163–177. Wiesbaden: Reichert.
- Paul, Ludwig. 2008. Kurdish Language, i. History of the Kurdish language. *Encyclopædia Iranica*. New York: Center for Iranian Studies, Columbia University. Online at: <http://www.iranica-online.org/articles/kurdish-language-i>, accessed 2 December, 2019.

- Ringe, Don A., Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1). 59–129.  
<https://doi.org/10.1111/1467-968X.00091>
- Sabethematabadi, Parisa, Erik Anonby & Jaffer Sheyholislami. 2017. A comparative analysis of language classification taxonomies in Kurdish, Tehrani, and Abade'i language communities. 3rd International Conference on Kurdish Studies (ICKS3), University of Exeter, June 26–28.
- Schleicher, August. 1853. Die ersten Spaltungen des indogermanischen Urvolkes [The first division of the early Indo-European people]. *Allgemeine Monatsschrift für Wissenschaft und Literatur* [General Monthly Journal for Science and Literature] 1853. 786–787.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen* [Relationships of the Indo-European languages]. Weimar: Böhlau.
- Schmitt, Rüdiger (ed.). 1989. *Compendium linguarum iranicarum* [Compendium of Iranian languages]. Wiesbaden: Reichert.
- Schmitt, Rüdiger. 2000. *Die iranischen Sprachen in Geschichte und Gegenwart* [The Iranian languages in history and the present]. Wiesbaden: Reichert.
- SJS = Sāzemān-e Jahād Sāzandegi. 1986. *Census of rural districts [dehestān] of Iran*. Tehran: Sāzemān-e Jahād Sāzandegi.
- Stilo, Don. 1981. The Tati language group in the sociolinguistic context of Northwestern Iran and Transcaucasia. *Iranian Studies* 14(3/4). 137–187. <https://doi.org/10.1080/00210868108701585>
- Stilo, Don. 2007. Isfahān, xxi. Provincial dialects. *Encyclopædia Iranica* 14(1). 93–112. New York: Center for Iranian Studies, Columbia University. Online at: <http://www.iranicaonline.org/articles/isfahan-xxi-provincial-dialects>, accessed 2 December, 2019.
- Stilo, Don. 2018. Caspian and Tatic. In Geoffrey Khan & Geoffrey Haig (eds.), *Language contact and language change in West Asia*, 659–824. Berlin: de Gruyter.
- Taheri-Ardali, Mortaza et al. 2016. Chahar Mahal va Bakhtiari Province. In: Anonby, Erik, Mortaza Taheri-Ardali et al. 2015–2020. *Atlas of the Languages of Iran (ALI)*. Ottawa: Geomatics and Cartographic Research Centre, Carleton University. Online at: [http://iranatlas.net/module/language-distribution.chahar\\_mahal\\_va\\_bakhtiari](http://iranatlas.net/module/language-distribution.chahar_mahal_va_bakhtiari)
- Taheri-Ardali, Mortaza, Erik Anonby, Amos Hayes. et al. Forthcoming. Atlas-e barkhatt-e zabānhā-ye Irān: Raveshshenāsi, tarāhi va natāyej-e avvalieh [The online *Atlas of the Languages of Iran*: Design, methodology and initial results]. *Jostārkhā-ye zabāni* [Language related research].
- TAVO = Orywal, Erwin (ed.). 1988. Karte 10, Vorderer Orient: Sprachen und Dialekte [Map 10, Middle East: Languages and dialects]. *Tübinger Atlas des Vorderen Orients (TAVO)*, series A, vol. 8. Wiesbaden: Reichert.
- Tedesco, Paul. 1921. Dialektologie der mitteliranischen Turfantexte [Dialectology of the Middle Iranian Turfan texts]. *Monde Oriental* 15. 184–258.
- Thomason, Sarah & Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- van der Wal Anonby, Christina. 2015. A grammar of Kumzari: A mixed Perso-Arabian language of Oman. Leiden: Leiden University dissertation.
- Windfuhr, Gernot. 1989. Western Iranian dialects. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum* [Compendium of Iranian languages], 294–295. Wiesbaden: Reichert.
- Windfuhr, Gernot. 1995. Dialectology. *Encyclopædia Iranica* 7(4). 362–370. New York: Center for Iranian Studies, Columbia University. Online at: <http://www.iranicaonline.org/articles/dialectology>
- Windfuhr, Gernot (ed.). 2009. *The Iranian languages*. London: Routledge.

## The additive particle in Persian

### A case of morphological homophony between syntax and pragmatics

Jila Ghomeshi

University of Manitoba

This chapter discusses the syntactic and semantic properties of the additive marker *-am* in Persian. I show that *-am* exhibits positional variability, is polysemous in meaning, and does not always contribute meaning that affects the truth conditions of the sentence. On the basis of this I classify *-am* as a pragmatic particle. Noting that *-am* is homophonous with both the first person singular agreement suffix and the first person pronominal enclitic, I propose that it is precisely because additive *-am* is a pragmatic particle that it is distinguished from the inflectional morphemes it resembles in form. Thus the chapter argues for three levels at which morphemes can be classified: derivation, inflection, pragmatic, and suggests that cross-level homophony is not accidental. Rather, frequency of use at one level predisposes a particular form to be used at another level. This ultimately gives a language its morphological ‘flavour’.

**Keywords:** additive, inflection, derivation, discourse particles, homophony, Persian

#### 1. Introduction

This chapter starts with the observation that morphological homophony is not rare, certainly where inflectional affixes and clitics are concerned, and possibly for other types of morphological processes as well. Through a detailed exploration of the additive marker in modern standard conversational Persian, I propose that some instances of homophony are illusory if we take morphological levels into account. Moreover, I propose that there are three such levels: inflection, derivation, and what I call pragmatics or the post-syntactic level for lack of a better term.

To illustrate, let’s consider English inflectional morphology as introduced in introductory linguistics textbooks. English is presented as having eight inflectional



affixes, one of which is the comparative *-er* which attaches to adjectives (O'Grady & Archibald 2016: 116). What is not usually mentioned in this context is that the comparative suffix is homophonous with the highly productive nominalizing suffix *-er* which derives agentive nouns from verbs:

- (1) a. comparative adjectives  
*small smaller*  
*blue bluer*  
*short shorter*
- b. deverbial agentive nominals  
*work worker*  
*run runner*  
*write writer*

In this particular case, the homophonous suffixes can be distinguished by the category to which they attach (adjectives vs. verbs), and more importantly by whether or not they change the lexical category of the base. The comparative suffix attaches to adjectives which remain adjectives while the nominalizing suffix attaches to verbs and creates nouns. It is this property that distinguishes them as inflectional and derivational, respectively – a distinction that complicates the issue of whether this is a case of homophony. If homophones are identical in form but different in meaning, where does the distinction between inflection vs. derivation fit in?

It is worth noting that the proper theoretical treatment of inflection vs. derivation remains elusive. For instance, inflectional morphemes can be characterized as syntactic based on the fact that they are productive, relatively exceptionless, 'grammatical', and can attach to phrasal constituents. In contrast, derivational morphemes are considered 'lexical' because they can be category-changing and more idiosyncratic in terms of their phonological effects on the base to which they attach. However, there are analyses in which category changing morphemes are syntactic (e.g. Baker & Vinokurova 2009; Kornfilt & Whitman 2011 for nominalizers), and theories in which there is no word-based lexicon that feeds syntax, rendering all morphology syntactic (e.g. Distributed Morphology, Halle & Marantz 1993; Embick & Noyer 2007).

In this chapter I argue that levels of morphology are important, regardless of how this is theoretically implemented. Assuming that true homophony only holds at a single level, I argue that frequency of use at one level predisposes morphemes for use on another level thus proliferating phonological identity but maintaining distinctions in use.

The chapter is organized as follows. In § 2, the phenomenon of homophony between inflectional and derivational morphemes is explored with examples from

both Persian and English. Section 3 discusses discourse markers and discourse particles, which serves to introduce the kinds of elements that are not part of the syntax proper. Section 4 presents a description of the syntax and semantics of the additive marker *-am*, on the basis of which I argue that, like other discourse particles, it is post-syntactic. In the conclusion I discuss the implications of this view of the grammar for the analysis of homophonous morphemes in general.

## 2. Inflectional and derivational homophony in Persian

Morphological homophony abounds in Persian. For instance, the *Ezafe* vowel in Standard Persian, *-e*, links nouns to their modifiers and possessors and can appear multiple times in the same noun phrase:<sup>1,2</sup>

- (2) *kif-e*      *čarm-e*      *siyāh-e*  
 bag-EZ      leather-EZ      black-EZ  
 head noun nominal modifier adjectival modifier  
*por az pul-e*      *in mard*  
 full of money-EZ      this man  
 phrasal adjectival modifier possessive nominal  
 “this man’s black leather bag full of money”

(adapted from Ghaniabadi 2009: 27 Example 27b)

Standard Persian does not have an independent word that corresponds to the English definite article *the*. In conversational spoken Persian, however, there is a suffix, also *-e* but stress-attracting, that appears on unmodified nouns to signal definiteness:<sup>3</sup>

1. The literature on the *Ezafe* construction is plentiful. Much of the recent work is discussed in Samiiian & Larson (this volume), who argue for a case-marking analysis. See also Ghomeshi (1997) and Ghaniabadi (2009) for a linker analysis.

2. The following abbreviations are used in the Persian examples: ADD = additive, CLC = clitic, CONT = continuous, COMP = comparative, COP = copula, DEF = definite, EZ = ezafe vowel, IMP = imperative, INDEF = indefinite, INTER = interjection, NEG = negative, PART = participle PL = plural, POSS = possessive, PRT = particle, SG = singular, SUBJ = subjunctive. Where examples are taken from other sources the transcription may be modified to fit the system used in this chapter.

3. This morpheme is discussed in Ghomeshi (2003, § 6) and Ghaniabadi (2009), however, Jasbi (this volume) takes a closer look at its semantics. Jasbi claims that in addition to definiteness, *é* introduces a uniqueness implication and thus can occur with indefinites making them scopally inert. I will continue to gloss *é* as DEF in spite of its more complicated semantics.

- (3) a. *kif*  
 bag  
 “bags/bag/the bag”  
 b. *kif-é*  
 bag-DEF.SG  
 “the bag”

*Ezafe* *-e* and definite *-é* are distributionally distinct: the former links nouns to modifiers WITHIN the nominal phrase but never at the end of a phrase while the latter is phrase-final, albeit when the phrase consists of the noun and at most one modifier.<sup>4</sup> The two suffixes are also phonologically distinct in that the *Ezafe* vowel is unstressed and definite *-é* is stressed:

- (4) a. *kif-e síyāh*  
 bag-EZ black  
 “black bag(s)”  
 b. *kif síyāh-é*  
 bag black-DEF.SG  
 “the black bag”

The stress difference between the *Ezafe* *-e* and the definite/uniqueness marker *-é* falls in line with the general distinction between inflectional and derivational morphology in Persian. For instance, agreement markers, the direct object marker *-rā*,<sup>5</sup> and the indefinite enclitic do not attract stress as shown in (5), while category-changing morphemes do, shown in (6)–(8) below:<sup>6</sup>

- (5) N + clitic possessor  
 a. *māšín* “car” *māšín-am* “my car”  
 car-1SG.POSS  
 N + object marker  
 b. *mórq* “chicken” *mórq-o* “the chicken”  
 chicken-OM  
 N + indefinite enclitic  
 c. *portoqál* “orange” *portoqál-i* “an orange”  
 orange-INDEF

4. When the noun and its modifier appear with the definite suffix *-é* the *Ezafe* vowel is not possible. See references cited above for more details on this point.

5. See Jasbi (this volume) for a discussion of the semantics of *-rā* and Karimi & Smith (also this volume), for an analysis of *-rā* as an accusative case marker within a theory of dependent case.

6. These examples are taken from Lazard (1992, § 237)

- (6) N → Adv
- a. *ettefāq* “chance” *ettefāq-án* “by chance”  
 b. *telefon* “telephone” *telefon-án* “by telephone”
- (7) N → Adj
- a. *xášm* “anger” *xášm-nák* “angry, furious”  
 b. *társ* “fear” *társ-nák* “terrifying”
- (8) V → N
- a. *xāstan/xāh* “to want” *xāh-és* “request”  
 b. *raftan/rav* “to go” *rav-és* “method”

This use of stress as a diagnostic for inflection vs. derivation in Persian yields surprising results when it comes to some non-category changing morphemes that have traditionally been considered inflectional in other languages. For example, number marking has been argued to be derivational (Ghomeshi 2003; Kahnemuyipour 2000) based on the fact that the plural marker *-ha* attracts stress. Kahnemuyipour (2019) using the same stress diagnostic argues that the comparative marker *-tar*, too, is derivational. As noted above, it is not clear what this means in a non-lexicalist model of grammar. For Kahnemuyipour, an inflectional morpheme instantiates a functional head in the syntax which leads him to propose that the comparative *-tar* is a non-projecting adjunct to an adjectival category-defining head.

Perhaps the best illustration of the work that a meaningful distinction between inflection and derivation can do comes from considering the uses of the suffix *-i*. Ignoring stress placement for a moment, this enclitic can attach to the end of nominal phrases to mark them as indefinite (9a), to nouns modified by relative clauses (9b), to adjectives and nouns to form abstract nouns (9c)–(9d), and to nouns to form adjectives (9e)–(9f), to nouns of time and adjectives to form adverbials (9g), (see Lazard 1992 for additional uses):

- (9) a. *ketáb* “book” *ketáb-i* “a book”  
 b. *ketáb* “book” *ketáb-i ke ...* “the book that ...”  
 c. *bozórg* “big” *bozorg-í* “bigness”  
 d. *márd* “man” *mard-í* “manliness”  
 e. *Irán* “Iran” *Irán-i* “Iranian”  
 f. *limú* “lemon” *limu-í* “which resembles a lemon”  
 g. *tanhá* “alone” *tanhā-í* “solitary”

Given the list above, to call *-i* a polysemous morpheme would miss a significant point: when this suffix has a truly grammatical function (that of marking (in) definiteness) it does not affect stress placement on the word to which it attaches. However, to leave it at that also misses a point which is that there are many different functions associated with one phonological form. That is, like *-e*, *-i* has a variety of functions in Persian on both the inflectional and derivational side:

Table 1. Inflectional vs. derivational *e* & *i* in Persian

Level	Form	Function	Form	Function
inflection	-e	<i>Ezafe</i>	-i	indefinite, relativizer
derivation	-é	definiteness	-í	forms abstract nouns, adjectives, and adverbials

While the stress-attracting properties of derivational *-é/-í* in Persian may be grounds for considering them phonologically distinct from their inflectional counterparts, and therefore not true homophones, English provides a similar example of a frequently used inflectional suffix that has taken on new life as a derivational morpheme. Note first that of the eight inflectional suffixes in English referred to in the introduction above, three of them are *-s*:

- (10) The English inflectional affixes (O'Grady & Archibald 2016: 116)
- ed past tense
  - en/-ed past participle
  - ing progressive
  - er comparative
  - est superlative
  - s third person singular non-past
  - 's possessive (genitive)
  - s plural

While this degree of homophony may be confusing enough, there appears to be a new use of *-s* that has arisen on social media where it marks words that have been clipped:

- (11) *totally* → *totes*  
*adorable* → *adorbs*  
*whatever* → *whatevs*  
*people* → *peeps*  
*feelings* → *feels*

It is unclear what the source for this innovative use is – perhaps the diminutive found with clipped proper names (*Jaiden* ~ *Jades*, *Trottier* ~ *Trotts*, *Kiana* ~ *Keeks*) or the internet play language LOLcats in which the plural marker is overextended, often with irregular orthography (Gawne & Vaughan 2011). In any case, we find an already overworked suffix acquiring additional duties.<sup>7</sup>

7. In this case the affix in question is simply a segment and it has been noted that unmarked segments can serve many morphological functions, not only in English but in German and Spanish (Janda 1987). I thank Joe Salmons for bringing Janda's work to my attention. The case of

Table 2. Inflectional vs. derivational *s* in English

Level	Form	Function
inflection	-s	plural, possessive, third person singular present
derivation	-s	diminutive, marking end of clipped words

The main question addressed in this chapter is: what if homophony between inflectional and derivational morphemes is not an accident but a fact of grammar? To be more specific, let us hypothesize that the productivity and/or frequency of use of a morpheme at one level (in this case inflection or derivation) makes it more likely to share its phonological form with a morpheme at another level. This principle is the grammatical analog of saying: if you want something done, ask a busy person to do it.

The proliferation of uses for a single affixal vowel or consonant of the sort we have seen above may be a challenge for theorists, but does not seem to pose a challenge for speakers. This is precisely the observation that Taylor (2003: 647) makes about polysemy in general, namely that it may raise theoretical and methodological issues for semanticists and practical issues for those working in natural language processing, for instance, but that speakers rarely experience it to be a problem at all. This leads to two observations:

- I. If phonologically identical affixes are distinguished by the *level* at which they attach, that is, if this is part of their morphological makeup, the instances of true homophony are reduced.
- II. A prolific form at one morphological level is *more* likely to correspond to a morpheme with a distinct function at another level by the principle of economy.

Building in the idea of morphological level in the representation of morphemes goes some way towards reducing the amount of homophony found in the grammar but in the following section I argue that a two-way distinction is not sufficient.

### 3. Discourse markers and discourse particles

Apart from the homophony between derivational and inflectional morphemes, there is also homophony between lexically or grammatically contentful morphemes on the one hand and discourse markers and particles on the other. I will use 'discourse marker' to refer to any element that has a textual function, i.e. something

---

Persian *-am* that I turn to in § 4 is different in that the affix is not made up of only one unmarked segment and yet the same proliferation of functions is exhibited.

that serves to sequence units of talk, and ‘discourse particle’<sup>8</sup> for any element that has an interpersonal function, i.e. that is dialogic (Traugott 2007, see also references cited therein). An exemplary case of a multi-functional element that has uses as a discourse marker and a discourse particle is *like* in English. D’Arcy (2017) identifies at least twelve different uses of ‘like’ seven of which are shown below:

- |      |    |  |                            |
|------|----|--|----------------------------|
| (12) | a. | <i>childlike</i>                                 | derivational suffix        |
|      | b. | <i>I like beans.</i>                             | verb                       |
|      | c. | <i>She grows vegetables and the like.</i>        | noun                       |
|      | d. | <i>That sounds like a good movie.</i>            | preposition                |
|      | e. | <i>I feel like you don’t care.</i>               | comparative complementizer |
|      | f. | <i>Like, I don’t know what you mean.</i>         | discourse marker           |
|      | g. | <i>The elephant is like just standing there.</i> | discourse particle         |

Of significance for our purposes are the use of *like* as a discourse marker and a discourse particle, shown in (12f) and (12g). D’Arcy defines discourse markers as clause-initial forms which “evaluate the relation of the current utterance to prior discourse” (2017: 57), whereas particles whose functions are primarily interpersonal and intersubjective can appear at multiple sites within an utterance. Through a detailed diachronic analysis she shows that the use of *like* as a discourse marker pre-dates its use as a particle. In other words, the use of *like* in a fixed position, as a clause-initial marker, pre-dates its use as a syntactically variable particle. In addition to its variable syntax, the meaning of the particle *like* is also variable. It can contribute meaning such as hedging, non-equivalence, approximation, or focusing (D’Arcy 2017: 15). I propose that these are precisely the properties that define post-syntactic elements: positional variability, polysemy, and non-truth-conditional contributions to meaning.

In Persian, a comparable example of homophony between a grammatical element and a discourse particle involves the complementizer *ke*. As a complementizer, *ke* is multifunctional and can introduce a purpose clause (13a), or a clause with a temporal (13b) or causal (13c) reading (these terms and categories are from Perry 2007). It can also introduce direct or indirect discourse (13d) & (13e) and an indicative complement clause (13f):<sup>9</sup>

---

8. Discourse particles are also called modal particles (e.g. Aijmer 1997; Traugott 2007) as they are epistemic, i.e. can signal subjective information on the part of the speaker. I follow D’Arcy (2017) in using discourse particle instead.

9. Estaji (2011) [1389] traces the sources of some of these uses of *ke* to relative pronouns and some to other connectives. See also Stilo (2004) for pronominal sources of *ke*.

- (13) a. *un manzel-o foruxt-an [(ke) be-r-an āmrikā]*  
 that house-OM sell.PAST-3PL that SUBJ-go-3PL America  
 “They sold that house [in order/so that] to go to America.”
- b. *hanuz vāred=na-shode bud-im [ke]*  
 yet enter= NEG-become.PASTPART COP.PAST-1PL that  
*mā-rā did*  
 1PL-OM see.PAST.3SG  
 “We had not yet entered when he saw us.” (Perry 2007: 996)
- c. *bo-ro birun [ke sob šod]*  
 IMP-go outside that morning become.PAST.3SG  
 “Go out, for it is morning.” (Perry 2007: 996)
- d. *goft [ke man ne-miyā-m]*  
 say.PAST.3SG that 1SG NEG-come-1SG  
 “He said ‘I’m not coming.’” (Perry 2007: 996)
- e. *goft [ke ne-miyā-d]*  
 say.PAST.3SG that NEG-come-3SG  
 “He said (he’s) not coming.” (Perry 2007: 996)
- f. *mi-dun-am [ke āftāb dāq-e]*  
 CONT-know-1SG that sun hot-COP-3SG  
 “I know that the sun is hot.”

While *ke* as a complementizer is syntactically associated with the following clause, it is phonologically cliticized to the matrix verb that precedes it. In some cases it is part of a larger string that functions as a single grammatical item, having undergone lexicalization and in some cases univerbation in Trousdale’s (2008) sense:

- (14) *ba’d az in ke* “after” [lit. “after that which”]  
*chun ke* “because”  
*barā-ye in ke* “for, because” [lit. “for that which”]  
*tā ke* “so that”  
*agar ke* “although”  
*balke* “but”

The status of *ke* as a complementizer has been questioned when it appears in sentences that are more controversially biclausal. For instance, it can appear following a modal verb as shown in (15a) below leading to a debate about whether it is or is not a complementizer in this case (see Ghomeshi 2001; the response by Darzi 2008, and references therein; see also Stilo 2004 who calls *ke* a subordinating particle). These cases aside, there is agreement that *ke* functions as a discourse particle in clearly mono-clausal sentences (see (15b) where it has been argued to be a focus marker or marker of emphasis (Lazard 1992; Windfuhr 1979: 70; and references therein). It can occur more than once and can follow any constituent of the clause



(see Oroji & Rezaei 2013 whose corpus study confirms that *ke* can appear anywhere but that also reveals a preference for *ke* to appear after subjects).

- (15) a. *Simā bāyad=ke be-re*  
 Sima should=?? SUBJ-go.3SG  
 “Sima should go.”
- b. *Simā=ke mi-re=ke*  
 Sima=KE CONT-go.3SG=KE  
 “SIMA is going (isn’t she?)”

In terms of its contribution to meaning *ke* as a discourse particle has been characterized as requesting confirmation, underlining the obvious (in exclamations), marking assertion in the context of counter-expectations (i.e. adversative), and identifying the MOST likely alternative from a list, i.e. having a scalar reading (see Bateni 2010; Lazard 1992; Oroji & Rezaei 2013).

Ghomeshi (2013) proposes that the multifunctionality of *ke* is in part explained by its having undergone a process of pragmaticalization (Erman & Kostinas 1993; Aijmer 1997). Pragmaticalization is similar to grammaticalization in that both processes involve a change over time that results in a lexical item taking on newer functions. In both cases the older uses can co-exist with the newer ones leading to ‘layering’ (Hopper 1991: 23; Hopper & Traugott 2003: 124–6) or heterosemy (Lichtenberk 1991; Diewald 2013). However, whereas grammaticalization leads to the development of functional words and morphemes (e.g. agreement affixes from pronouns, auxiliary verbs from main verbs), pragmaticalization produces discourse markers and particles. In accounting for the changes over time exhibited by lexical items, van Gelderen (2011) suggests that there are several principles involved. For example, the change from pronoun to clitic to agreement or from relative pronoun to complementizer can be formalized in terms of a preference for lexical items that occur in phrasal positions to be reanalyzed as heads:

(16) **Head Preference Principle (HPP)**

Be a head, rather than a phrase. (van Gelderen 2004, 2011: 13 Example 15)

The tendency for elements to merge at higher and higher points in a syntactic structure as in the case of the development of auxiliary verbs from main verbs is captured by the following principle (cf. Chomsky’s 1995, 2001 ‘merge-over-move’ principle; see also Roberts & Roussou 1999):

(17) **Last Merge Principle**

Merge as late as possible. (van Gelderen 2011: 14 Example 17)

Ghomeshi (2013) notes that the Head Preference Principle does not explain the development of the conversational particle *ke* from the complementizer as both are

head-like. Nor does the Last Merge Principle fare any better as the complementizer *ke* is arguably merged as the highest head in the left periphery of the clause. In order to account for the fact that *ke* seems to occur on a number of different constituents in a clause, Ghomeshi proposes that *ke* has lost its selectional features – those that are involved in determining head-complement relations – as well as its categorial status<sup>10</sup> by a principle of Detachment:

(18) **Detachment Principle**

[head complement] > [<sub>adjunct</sub> head complement]

(Ghomeshi 2013: 9 Example 13)

Principles such as the Head Preference Principle, the Late Merge Principle and the Detachment Principle account for the change that a linguistic unit can undergo over time. Taking a diachronic view does not fully account for the distribution of the particle *ke*, however. The syntax of pragmatization involves the three properties, according to Ghomeshi (2013): (i) a lexical item loses its ability to take a complement, (ii) a lexical item becomes category-less, and (iii) a lexical item becomes syntactically mobile.<sup>11</sup> It is the last property that is the most challenging to account for within a formal syntactic framework, where head-complement relations and a dedicated position with extended functional structure are expected. Recall that *ke* is always enclitic but as a complementizer is associated with the FOLLOWING constituent (see 13a) while as a conversational particle it has scope over the constituent that PRECEDES it, as shown below.<sup>12</sup>

- (19) a. *Simā=ke bā mā miy-ād*  
 Sima=KE with us CONT-come.3SG  
 “Sima (at least) is coming with us.”

---

10. Decategorization or decategorialization (both terms are used in the literature) is one of the effects of grammaticalization (Hopper 1991). It is not implausible to presume it holds of pragmatization as well.

11. These properties have been identified in the change in English of ‘complement taking mental predicates’ such as *I think*, *I suppose*, into adverbial-type expressions with epistemic meanings (see Van Bogaert 2011, for instance). Thus, *know* which can take a nominal or clausal complement, can also serve as an adverbial in combination with *you*:

- i. *You know the answer.*
- ii. *You know that cars are expensive.*
- iii. *(y’know) cars are (y’know) expensive (y’know)*

12. The translations are approximate. As has been noted about modal particles (here called discourse particles), they are difficult to translate and a true investigation of their contribution in meaning requires corpus and/or experimental study.

- b. *Simā bā mā=ke miy-ād*  
 Sima with us=KE CONT-come.3SG  
 “Sima is coming with US.”
- c. *Simā bā mā miy-ād=ke*  
 Sima with us CONT-come.3SG=KE  
 “Sima is coming with us (isn’t she?)”

While a full theory of the syntax of discourse particles remains to be developed, the point here is that like English *like*, Persian *ke* exhibits a set of properties that make it unlikely to correspond to a dedicated functional projection such as FocP (Focus Phrase): it has positional variability, is polysemous and makes a non-truth-conditional contribution to the meaning of the proposition. These are the properties that characterize a pragmatic rather than a syntactic element. In the next section we turn to the additive marker which we will see similarly fits the same criteria for being a post-syntactic element.

#### 4. The additive marker in conversational Persian

Stilo (2004) in his overview of coordination three western Iranian languages, Vafsi, Persian and Gilaki, presents a number of different coordination strategies in Persian one of which is related to the additive marker. By way of contrast, we start with the most common type of coordination. Persian, like English, has simple coordinate conjunction using an enclitic *-o* which can connect two or more constituents of the same type:

- (20) a. *sib=o berenj(=o gusht=o ...)* *xarid-am*  
 apple=CONJ rice(=CONJ meat=CONJ ...) buy.PST-1SG  
 “I bought apples and rice (and meat and ...).”
- b. *sib xarid-am=o berenj*  
 apple buy. PST-1SG=CONJ rice  
 “I bought apples and rice.”

Example (20b) above shows that constituents connected through conjunction can be discontinuous and when they are separated, the conjunction *-o* appears with the second conjunct even though it is phonologically dependent on the immediately preceding element, in this case the verb (see Stilo 2004: 280, who refers to this as ‘extraposition’).<sup>13</sup>

13. We see this difference between syntactic constituency and phonological behaviour in the case of *ke* as well where as a complementizer it is syntactically a constituent with a subordinate clause but is phonologically dependent on the matrix verb or whatever else immediately precedes it.

Persian also has bisyndetic coordination<sup>14</sup> involving the coordinators *ham ... ham ...* “both ... and ...”:

- (21) a. *ham sib(=o) ham berenj xarid-am*  
 CONJ apple(=CONJ) CONJ rice buy. PST-1SG  
 “I bought (both) apples and rice.”  
 b. *ham sib xarid-am(=o) ham berenj*  
 CONJ apple buy. PST-1SG(=CONJ) CONJ rice  
 “I bought (both) apples and rice.”

Stilo (2004: 273) traces *ham* back to Old Persian *ham* meaning “together, with” and *hama* meaning “one and the same” and, ultimately, to the Proto-Indo-European form that also gives English *same*. As with *ke*, *ham* participates in compounds as shown in (22) (all examples taken from Stilo’s discussion of *ham* as a derivational morpheme (2004: 324–326)):

- (22) a. *ham vatan* HAM-homeland “compatriot”  
 b. *ham kelās(-i)* HAM-class- ‘ite’ “classmate”  
 c. *ham sāye* HAM-shade “neighbour”  
 d. *ham diġe* HAM-other “each other”  
 e. *bā ham* with-HAM “together”  
 f. *hamin* HAM-this “this very one”  
 g. *hamiše* “always”

Stilo does not trace the connection between coordinator *ham* and what he calls the inclusive focus particle *-am*, referred to as the additive marker here. He presents the particle use as simply another use of *ham* (Stilo 2004: 323–324) in all three of the languages that he considers, Vafsi, Persian, and Gilaki. I will likewise assume that the two are etymologically related, something that is reinforced by the fact that they are written the same way in Persian. There are nevertheless significant differences between them. The additive marker is not a coordinator, it does not join two constituents of the same type but takes scope over a proposition. Thus a sentence containing the bisyndetic coordinator *ham ... ham ...* has to be repeated if the additive is used instead:

---

14. Haspelmath (2004: 4) defines ‘bisyndetic coordination’ as involving one coordinator per coordinand. It contrasts with ‘monosyndetic coordination’ where two coordinands appear with just one coordinator. Both types of coordination can involve more than two coordinands in which case bisyndetic has an equal number of coordinators to coordinands while monosyndetic coordination has one fewer.

- (23) a. *ham sib(=o) ham berenj xarid-am*  
 CONJ apple(=CONJ) CONJ rice buy.PST-1SG  
 “I bought (both) apples and rice.”  
 b. *sib xarid-am, berenj-am \*(xarid-am)*  
 apple buy. PST-1SG rice-ADD buy.PST-1SG  
 “I bought apples (and) I also bought rice/I bought rice too.”

Karvovskaya (2013: 80) in her discussion of the additive marker *-məs* in Ishkashimi provides a test for additivity that she adopts from Berger & Höhle (2012). Consider a situation in which a mother leaves her child with an apple and an apricot and goes away for a while. Upon her return she asks ‘Did you eat the apple?’ The answer in (24a) below with an additive marker is appropriate if the child has eaten *both* the apple and the apricot. Neither a simple coordinator like *va* ‘and’ nor the bisyndetic coordinator *ham* is felicitous in this context (24b):

- (24) a. *zard-ālu-r-am xord-am*  
 yellow-plum-OM-ADD eat.PST-1SG  
 “I ate the apricot too.”  
 b. *\*va/ham zard-ālu-ro xord-am*  
 and/both yellow-plum-OM eat.PST-1SG  
 #“And I ate the apricot.”

Another difference between the additive marker and the bisyndetic coordinator *ham ... ham ...* is that the latter can link verbs. The incompatibility between the additive *-am* and a verb is something that is characteristic of additive markers cross-linguistically (Forker 2016: 4):<sup>15</sup>

- (25) *berenj-o ham xarid-am=o ham poxt-am*  
 rice-OM CONJ buy. PST-1SG=CONJ CONJ cook. PST-1SG  
 “I (both) bought the rice and cooked (it).”  
*\*berenj-o xarid-am poxt-am-am*  
 rice-OM buy. PST-1SG cook. PST-1SG-ADD  
 “I bought the rice and cooked (it).”

In the next section I will provide a review of the syntactic distribution of *-am* followed by a description of its contribution to meaning and a discussion of its form.

---

15. We will see below that this is not due to a constraint on two *-am*'s in a row.

## 4.1 The syntactic distribution of the additive marker

The data in this and the following sections are all from the Callfriend Farsi corpus of telephone speech. The corpus was collected by the Linguistic Data Consortium (<https://www ldc.upenn.edu>) and released in 1996. It consists of 60 unscripted phone conversations between 5 and 30 minutes long. The callers are native Persian speakers who were in the United States or Canada at the time of the call. For this chapter, approximately 300 examples involving *-am* were gathered from 12 calls representing approximately 296 minutes of conversation. From these examples the following descriptive generalizations emerge.

There is a strong tendency for the additive marker to appear in second position which can either be after the first constituent, as shown in (26) or after the first word if the word constitutes a syntactic constituent, as shown in (27):

- (26) a. *bā āb-e namak-am bāyad galu-t moratab*  
 with water-EZ salt-ADD should throat-2SG.POSS constantly  
*bu-šur-i*  
 SBJ-wash-2SG  
 “you should also gargle [*lit.* “wash your throat”] constantly with salt water”  
 [FA 4054, 1:28]
- b. *un copy-ā-ye pāsport-am be-dard na-xord?!*  
 that copy-PL-EZ passport-ADD to-pain NEG-hit.past.3SG  
 “And those copies of the passport were useless too?!” [FA 4218, 1:23]
- (27) a. *fārsi-am bāyad harf=be-zan-i*  
 Farsi-ADD should talk=SBJ-hit-2SG  
 “and you have to speak Farsi” [FA 4054, 0:16]
- b. *dīšab-am xeyli sard bud*  
 last.night-ADD very cold COP.past.3SG  
 “last night was really cold too” [FA 4054, 0:16]

It can occur twice in an utterance and can occur in an embedded clause:

- (28) a. *māmān-am goft agar*  
 mother-ADD say.PAST.3SG if  
*davat-am-am=mi-kard-an ne-mi-raft-am*  
 invitation-1SG-ADD=CONT-do.PAST-3PL NEG-CONT-go.PAST-1SG  
 “and Mother said even if they’d invited me I wouldn’t go”  
 [FA 4054, 13:28]

- b. *bābā, bad az in telefon dobāre bet*  
 INTER after of this telephone again to-2SG.CLC  
*zang=mi-zan-am, [pause] ye zare-am be qoli az to-am*  
 ring=CONT-hit-1SG one bit-ADD to promise of you-ADD  
*harf=be-zan-im [joint laughter]*  
 talk=SUBJ-hit-1PL  
 “listen, after this call, I’ll call you again and let’s talk a little about you too.”  
 [FA 4218, 5:36]
- c. *fekr=kon-am zud-tar-am bar=gard-an emsāl*  
 think=do-1SG fast-COMP-ADD back=wander-3PL this.year  
 “I think they’ll return sooner this year too.” [FA 4117, 13:28]

It can appear after pronominal enclitics that can indicate possession, including the first person possessor pronominal enclitic with which *-am* is homophonous:<sup>16</sup>

- (29) *ye province-e digar-i injā hast be nām-e Alberta,*  
 one province-EZ other-INDEF here is to name-EZ Alberta,  
*markaz-eš-am Edmonton-e*  
 centre-3SG.POSS-ADD Edmonton= COP.PRES.3SG  
 “... there’s another province here called Alberta and its capital is Edmonton”  
 [FA 4383, 8:20]
- baʔd emruz sar-e kār-am-am goft-am,*  
 then today head-EZ work-1SG.POSS-ADD say.PAST.1SG  
*ticket-am-am order=dād-am*  
 work-1SG.POSS-ADD order=give.PAST-1SG  
 ‘[with list intonation] then today I told [them] at work and I ordered my ticket’  
 [FA 4219, 18:56]

It can also appear after *ham*:

- (30) [woman talking about Walmart; she starts by saying their shoes are not bad and her conversational partner agrees and says a few things, then she carries on as follows]
- kafš-ā-š ke soft-e, [vā in] ham*  
 shoe-PL.3SG.POSS PRT soft-COP.PRES.3SG [unintelligible] CONJ  
*rāhat-e, masalan mi-xā-m bache baqal=kon-am,*  
 comfortable-COP.PRES.3SG like CONT-want-1SG child hug=do-1SG  
 [pause] *ham-am xeyli arzun-e [pause] kafš-ā-š*  
 CONJ-ADD very cheap-COP.PRES.3SG shoe-PL.3SG.POSS  
 “their shoes are soft, ... (they’re) both comfortable, like when I want to carry a child, and they’re very cheap, their shoes” [FA 4300, 26:34]

16. Where English words are used in Persian conversation they will be underlined.

There are two constraints on additive *-am* that are possibly connected. The additive marker does not appear on verbs nor does it appear in clause-final position.<sup>17</sup> The two constraints are needed independently as verbs need not be clause-final (see (31a) below) and consequently other elements can appear at the end of a clause. The constraint on verbs is specific to inflected verbs,<sup>18</sup> i.e. verbs bearing subject agreement, as the additive can appear on modal verbs (see (31b)) and verbal nouns (31c) neither of which take agreement:

- (31) a. *umad-im -chiz- piš-e xānum Kalāli, fardā-m*  
 came.PAST-3PL -thing- beside-EZ Mrs. Kalali tomorrow-ADD  
*mi-r-im xune āzar xanum*  
 CONT-go-3PL house Azar lady  
 “We’ve come [to visit] Mrs. Kalali and tomorrow we’ll go to Mrs. Azar’s house.” [FA 4117, 0:29]
- b. *na-bāyad-am bo-kon-an*  
 NEG-should-ADD SUBJ-do-3PL  
 “and they shouldn’t/ nor should they” (ratifying something the other speaker has said) [FA 4219, 15:54]
- c. [a woman talking about her spouse or son and his pet snake]  
*na be-heš mi-res-e, tamiz=mi-kon-e qazāš-o*  
 no to-3SG CONT-arrive-3SG clean=CONT-do-3SG food-OM  
*mi-d-e, na vali xob vaqte in=ke bi-ād*  
 CONT-give-3SG no but well time that=CMPLZ SUBJ-come.3SG  
*be-šin-e bāh-āš bāzi=kon-e ... bāzi=kardan-am*  
 SUBJ-sit-3SG with-3SG play= CONT-do-3SG play=do-ADD  
*na-dār-e! āxe chi=kār mi-kon-e un !!*  
 NEG-have-3SG PRT what=work CONT-do-3SG that  
 “no, he takes care of it, cleans it, give it its food, no but well he doesn’t have the time to come sit and play with it, ... and playing isn’t even possible! [I mean] what does it even do?” [FA 4117, 22:08]

Finally, *-am* can also appear on the non-verbal element in a complex predicate (32a) and in idiomatic expressions (32b):

17. Out of the over 300 examples gathered, there are three cases of the additive that appear to be on an element in clause-final position. In two cases the last element is *un-am* where the additive is attached to a pronoun *un* “her/him” that refers back to the subject and in the third case it is on *vāqean-am* “really, seriously”. In all three cases there is a pause before the items in question so I don’t consider them to be part of the preceding clause.

18. According to Forker (2016: 4) there are a few languages that allow an additive to occur on a finite verb but this is not typical.





that are either extensions of meaning or diachronic developments. An example of a core additive use of Persian *-am* is given below in (34a). Stilo (2004: 323) refers to *-am* in this use as an inclusive focus particle. Forker notes that additives can be used in negative clauses in which case they are translated by ‘either’, but unlike ‘either’ they need not be polarity-sensitive in other languages. This is true of Persian *-am* which can occur in negative sentences in the sense of ‘either’ as shown in (34b) but is not polarity-sensitive:

- (34) *Razavi-am bud bāhā-šun?*  
 Razavi-ADD COP.PAST.3SG with-3PL  
 “Was Razavi with them too?” [FA 4344, 2:23]  
*unā-m ne-mi-ān dige ā?*  
 they-ADD NEG.COME.3PL then PRT  
 “They’re not coming either then huh?” [FA 4451, 15:44]

Related to the core additive is the scalar additive where the associate of the additive is the least likely among a set of alternatives for which the proposition holds. In negative clauses the scale is reversed and the associate is the most likely of the set of alternatives on the scale (Forker 2016: 73). Stilo (2004: 324) describes *-am* as having acquired a sense of ‘even’ and in this sense often occurs with *hattā* ‘even’ in the same clause, but need not as the examples below show:

- (35) *dah sāl-am be-mun-i hamin-e, harchi-am be-mun-i*  
 ten year-ADD SUBJ-stay.2SG same-COP.3SG whatever-ADD SUBJ-stay.2SG  
*hamin-e*  
 same-COP.3SG  
 “If you stay even ten years it’s the same, however long you stay it’s the same”  
 [FA 4399, 13:26]  
*dar amrika xarčang-am mi-xor-an*  
 in America crab-ADD CONT-eat.3PL  
 “In America they even eat crabs.” [Stilo 2004: 324 Example 191]

The third use of the additive marker in Persian corresponds to what Forker (2016, § 3.2) identifies as association with contrastive topics and topic switch, attested for around half of the 42 languages she investigates. Stilo (2004) calls this use of *-am* an adversative coordinator that means “and, but”:

- (36) *faqat mi-xāst ke pedar-eš bar=gard-e,*  
 only CONT+want.PAST.3SG that father+3SG.POSS PRT=turn+3SG  
*pedar-e-ham fekr-e bar=gašt-an na-dāšt*  
 father+DEF-ADD thought+EZ PRT=turn.PAST+INF NEG+have.PAST.3SG  
 “He only wanted his father to return. But the father had no thought of return-  
 ing.” [Stilo 2004: 324 Example 192]

Stilo mentions, however, that Lazard (1989: 281) considers the use of *ham* above to mark a new theme. This is consistent with the idea that the additive can mark contrastive or new topics. Forker (2016) notes for contrastive topics, that there are parallel or near parallel predications that hold of one element in the set of alternatives but not another. The adversative sense can arise when contrast involves rejection or correction (hence, negation). However, she notes that additives can also associate with non-contrastive ‘aboutness’ topics, which she takes to be a “grammaticalization from their use with contrastive topics” (2016: 76) In this case, a topic switch has taken place. This is perhaps the best analysis of a commonly occurring use of the additive *-am* in the Callfriend Corpus whereby the additive marker appears in answers to questions as the following exchanges show:

- (37) a. A: *māšin-et dorost=šod?*  
           car-3SG.POSS fix=become.PAST.3SG  
           “Did your car get fixed?”  
       B: *māšin-am-am na, hamun juri-e*  
           car-1SG.POSS-ADD no, same way-COP.3SG  
           “my car, no, it’s still the same” [FA 4054, 2:12]
- b. A: *xob bache-hā četor-an?*  
           well child-PL how-COP.3PL  
           “So how are the kids?”
- c. B: *bache-hā-m xub-an*  
           child-PL-ADD well-COP.3PL  
           “the kids are well” [FA 4117, 2:12]
- d. A: *Maria chi kār mi-kon-e?*  
           Maria what work CONT-do-3SG  
           “What’s Maria up to?”  
       B: *Maria-m xub-e, salām mi-resun-e*  
           Maria-ADD well-COP.3SG hello CONT-send-3SG  
           “Maria’s well, she says hi” [FA 6871, 4:33]

In the examples above, the additive marker appears on the constituent about which the question has been asked. There are at least as many instances of *-am* in this use as there are of *-am* within a question (e.g. in (34a) above) where it can be analyzed as a focus marker and therefore presuppositional. In these cases it is not identifying one of a contextually relative set of alternatives, but rather is marking a new topic.

The final function that the additive marker serves is as a conjunctive adverb (Forker 2016) or conjunctive coordinator (Stilo 2004) meaning ‘and’ or ‘and then’. Forker notes that in this function the additive connects sentences in narrative units and often attaches to a temporal adverbial:

- (38) a. *fardâ-am mi-r-im xune âzar xanum*  
 tomorrow-ADD CONT-go-1PL house Azar lady  
 “and tomorrow we’ll go to Azar Xanom’s house” [FA 4117, 0:29]
- b. *qablan-am tu arteš bud-e*  
 before-ADD in army COP.PAST-3SG  
 “and he was in the army before” [FA 4117, 20:58]
- c. *ba?d-eš-am mi-r-e Washington*  
 then-3SG.CLC-ADD CONT-go-3SG Washington  
 “and then she’ll go to Washington” [FA 4219, 21:05]
- d. *xeyli-am behtar az Caspian-e*  
 much-ADD better from Caspian-COP.3SG  
 “and it’s much better than Caspian [a café]” [FA 6690, 15:33]

The functions of the additive that Forker (2016) finds in some languages but that do not appear in Persian include a concessive sense (introducing clauses translated with ‘although’ ...), the formation of indefinite pronouns, and constituent coordination (also called ‘emphatic coordination’). In the case of this last function, Persian can use *ham ... ham ...* as bisyndetic coordination, but as discussed in the beginning of this section, the marker *-am* cannot be used in this way. Table 4 summarizes the uses of *-am* in Persian:

**Table 4.** The meaning of additive *-am*

core additive	✓
scalar additive	✓
contrastive topic and topic switch marker	✓
conjunctive adverb	✓
concessive	×
indefinite pronoun	×
emphatic coordination	×

The description of the meaning of additive *-am* in Persian in this section highlights the polysemy exhibited by this morpheme. This is independent of its positional variability that we saw in the previous section – something that König (1991) identifies as being characteristic of additives. It is important to note that differences in meaning do not correlate with different syntactic positions. The marker *-am* in second position in a Persian sentence can be functioning as a core additive (in the sense of ‘too/also’), as the marker of a new topic, or as a conjunctive adverb meaning ‘and then’. Finally we can note that *-am* makes a non-truth-conditional contribution to propositions in at least three out of four of its uses (excluding the core additive reading) so in sum meets the criteria for being considered a discourse

particle. In the next section we will turn to the phonological form of the additive marker showing that it is identical to two other frequently used suffixes in conversational Persian.

### 4.3 The form of the additive marker

Persian has two sets of agreement markers: one set marking subject agreement on verbs and the other, usually referred to as the set of pronominal enclitics, marking a variety of other functions.<sup>20</sup> Typical uses of the pronominal enclitics include their appearance on nouns as possessors, on prepositions as objects of those prepositions, and on the nonverbal element of complex predicates as direct objects as shown in (39):

- (39) a. *ketâb-eš jâleb-e*  
 book-3SG.POSS interesting-COP.3SG  
 “his/her book is interesting”  
 b. *b-eš dâd-am*  
 to-3SG.CLC give.PAST.1SG  
 “I gave [it] to him/her”  
 c. *davat-eš=kard-am*  
 invitation-3SG.CLC=do.PAST.1SG  
 “I invited him/her”

The paradigms for the two sets of agreement markers in their conversational rather than formal pronunciation are given in Table 5:

**Table 5.** Subject agreement suffixes and pronominal enclitics in Persian

	Subject agreement		Pronominal enclitics	
	SG	PL	SG	PL
first	-am	-im	-am	-emun
second	-i	-in	-et	-etun
third	∅	-an	-eš	-ešun

We see in the table above that the only place where the subject agreement suffixes and the pronominal enclitics have the same form is in the first person singular. That form is *-am*, which is identical to the additive. This means that it is possible to have sentences in conversational Persian in which almost every word ends in a different *-am*:

20. See Haig (this volume) for a diachronic look at these two sets of inflectional morphemes across Iranian languages.

- (40) *fekr=mi-kon-am*    *dišab-am*    *barādar-am*    *bilit-am-o*  
 thought=COP-do-1SG last night-ADD brother-1SG.POSS ticket-1SG.POSS-OM  
*barā-m*    *xarid*  
 for-1SG.POSS buy.PAST.3SG  
 “I think my brother bought my ticket for me last night too.”

It is important to emphasize that such sentences are not at all confusing to Persian speakers. As we saw in § 3 with *ke*, which has both a grammatical and pragmatic status, *-am* serves as both an inflectional morpheme and a pragmatic particle. As a pragmatic particle it exhibits syntactic variability, polysemy, and does not contribute to the truth-conditions of the proposition in which it occurs. The morphological distinction between agreement as inflectional and the additive marker as pragmatic means that phonological identity in this case does not entail true homophony.

## 5. Conclusion

In this chapter I have made a number of claims. First I have proposed that in addition to the two levels of morphology that we identify as derivational vs. inflectional there is a third level, the pragmatic level, at which discourse markers and particles combine with the output of syntax. While these elements may be independent words in some languages, we have evidence in Persian from the focus marker *ke* and the additive *-am* that they can also be clitics or affixes. We have seen that they can appear in different positions and can appear more than once per clause. They contribute meanings that may have to do with the sequencing of narrative texts or with interpersonal and epistemic meanings that reflect speaker stance. While such particles are well-studied by sociolinguists and researchers interested in grammaticalization, the proposal here has focused on their morphological representation. Specifically, I claim that part of the information we store along with sound and meaning is a morpheme’s status as derivational, inflectional or pragmatic. Two morphemes that have the same phonology but different meaning are not homonyms if they belong to different levels of the grammar.

The second claim that I have made is that productivity or frequency of use at one level predisposes a phonological form to take on meanings and functions at another level. In the case of the additive, this means that its identity in form with the frequently used first person singular agreement suffix and pronominal enclitic, is not merely accidental. Or, to go back to an example discussed at the beginning of § 2, the use of the vowel *-e* as the linking vowel in the *Ezāfe* construction and as a definiteness marker is similarly not a coincidence. In fact, given that it is a frequently used vowel, it is not surprising that the same vowel with rising intonation

can be used in utterance-initially in conversation by a speaker to indicate surprise. That is ‘*e ... ?*’ in Persian is used like ‘really ... ?’ in English.

**Table 6.** Types of *e* in Persian

Level	Form	Function
pragmatics	<i>e</i>	utterance-initial interjection indicating surprise
inflection	- <i>e</i>	Ezafe
derivation	- <i>é</i>	definiteness

These proposals are not specific to Persian. Thus in Korean, for instance, the plural marker has the form *-tul* and it has been noted that it has uses in which it does not seem to indicate plurality. Song (1997) refers to this phenomenon as ‘extrinsic plural marking’ and suggests that the plural marker is copied from the subject onto other constituents to mark ‘distributivity’. However, noting that there are uses of *-tul* that neither mark plurality nor distributivity, Song goes on to propose that it can also serve as a focus marker. He gives the following somewhat contrived by theoretically possible sentence to support this claim:

- (41) *ai-tul-i kongwon-eyse(-tul) chinkwu-lang(-tul) culkepkey(-tul)*  
 child-PL-NOM park-LOC(-PL) friend-COM(-PL) cheerfully(-PL)  
*nolay(-tul)-ul pwulu-ko(-tul) siph-e(-tul) ha-ess-ta*  
 song(-PL)-ACC sing-COM(-PL) like-CONJ(-PL) do-PST-IND  
 “The children wanted to sing a song cheerfully with their friends in the park.”  
 [Song 1997: 218 Example 38]

We can see that like the proliferation of *-am* in a Persian sentence, the phonological resemblance between the plural marker and accusative case marker in Korean, along with the fact that both can be used more than once in a sentence (see Schütze 2001 on case stacking in Korean and the use of accusative *-lul* as a focus marker) gives the conversational language its unique flavour. While the ideas put forward in this chapter have yet to be worked out in more detail, the facts suggest that working out the puzzle of morphological homophony will reveal significant insights into the structure of the grammar.

## Acknowledgements

I would like to thank the organizers of NACIL1, and in particular Richard Larson, for a truly memorable and stimulating conference. I would also like to gratefully acknowledge the additional work that the editors of this volume took on in service of the community of scholars who work on Iranian linguistics. Vida Samiian and Sedigheh (Sophie) Moradi were patient, prompt, and

unfailingly supportive in seeing each stage of my contribution through and the comments of the anonymous reviewers and of the executive editor, Joe Salmons, were very helpful. I would like to thank two research assistants, Sharareh Esmaeili and Marzieh Hadei, who gathered data for me over the time my interest in the additive marker took hold. Finally, I find it hard to convey the depth of my appreciation for my peers and students in this community of Iranian linguistics who inspire me and push my ideas forward. All errors are my own.

## References

- Aijmer, Karin. 1997. *I think* – an English modal particle. In Toril Swan & Olaf Jansen Westvik (eds.), *Modality in Germanic languages: Historical and comparative perspectives*, 1–47. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110889932.1>
- Baker, Mark C. & Nadya Vinokurova. 2009. On agent nominalizations and why they are not like event nominalizations. *Language* 85. 517–556. <https://doi.org/10.1353/lan.0.0144>
- Batani, Mohammad R. 2010. Application of *ke*, *dige*, *akhe*, *ha*. *Bukhara Arts and Culture Journal* 70. 38–51. Translated by Sharareh Esmaeili.
- Berger, Frauke & Barbara Höhle. 2012. Restrictions on addition: Children’s interpretation of the focus particles *auch* (also) and *nur* (only) in German. *Journal of Child Language* 39. 383–410. <https://doi.org/10.1017/S0305000911000122>
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.
- Darzi, Ali. 2008. On the vP analysis of Persian finite control constructions. *Linguistic Inquiry* 39. 103–116. <https://doi.org/10.1162/ling.2008.39.1.103>
- D’Arcy, Alexandra. 2017. *Discourse-pragmatic variation in context: Eight hundred years of LIKE*. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.187>
- Diewald, Gabriele. 2013. “*Same same but different*” – Modal particles, discourse markers and the art (and purpose) of categorization. In Liesbeth Degand, Bert Cornillie & Paola Pietrandrea (eds.), *Discourse markers and modal particles*, 19–45. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.234.02die>
- Embick, David & Rolf Noyer. 2007. Distributed morphology and the syntax morphology interface. In Gillian Ramchand & Kurt Rice (eds.), *The Oxford handbook of linguistic interfaces*, 289–324. Oxford: Oxford University Press.
- Erman, Britt & Ulla-Britt Kostinas. 1993. Pragmaticalization: The case of *ba’* and *you know*. *Studier i Moderspråkvetenskap* 10. 76–93.
- Estaji, Azam. 2011 [1389]. A historical study of homophonous *ke* in Persian. *Journal of Linguistics and Dialects of Khorasan* 2:3. 1–13. Translated by Sharareh Esmaeili.
- Forker, Diana. 2016. Toward a typology for additive markers. *Lingua* 180. 69–100. <https://doi.org/10.1016/j.lingua.2016.03.008>
- Gawne, Lauren & Vaughan, Jill. 2011. I can haz language play: The construction of language and identity in LOLspeak. In Maïa Ponsonnet, Loan Dao & Margit Bowler (eds.), *Proceedings of the 42nd Australian Linguistic Society Conference, Canberra ACT, 2–4 December 2011*, 97–122, ANU Research Repository. <http://hdl.handle.net/1885/9398> (12 June, 2017).
- van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.71>



- van Gelderen, Elly. 2011. *The linguistic cycle*. New York: Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780199756056.001.0001>
- Ghaniabadi, Saeed. 2009. The empty noun construction in Persian. Winnipeg, MB: University of Manitoba dissertation.
- Ghomeshi, Jila. 1997. Non-projecting nouns and the *ezafe* construction in Persian. *Natural Language and Linguistic Theory* 15. 729–788. <https://doi.org/10.1023/A:1005886709040>
- Ghomeshi, Jila. 2001. Control and Thematic Agreement. *Canadian Journal of Linguistics* 46. 9–40.  
<https://doi.org/10.1017/S0008413100017928>
- Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia Linguistica* 57. 47–74. <https://doi.org/10.1111/1467-9582.00099>
- Ghomeshi, Jila. 2013. The syntax of pragmaticalization. In Shan Luo (ed.), *Actes du congrès annuel de l'Association canadienne de linguistique 2013. Proceedings of the 2013 annual conference of the Canadian Linguistic Association*. <http://cla-acl.ca/actes-2013-proceedings/> (19 November, 2013).
- Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth L. Hale & Samuel Jay Keyser (eds.), *The view from building 20. Essays in honor of Sylvain Bromberger*, 111–76. Cambridge, MA: MIT Press.
- Haspelmath, Martin. 2004. *Coordinating constructions*. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/tsl.58>
- Hopper, Paul J. 1991. On some principles of grammaticalization. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, 17–35. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/tsl.19.1.04hop>
- Hopper, Paul J., & Traugott, Elizabeth C. 2003. *Grammaticalization*, 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139165525>
- Janda, Richard David. 1987. On the motivation for an evolutionary typology of sound-structural rules. Los Angeles, CA: University of California, Los Angeles dissertation.
- Kahnemuyipour, Arsalan. 2000. On the derivationality of some inflectional affixes in Persian. Paper presented at the annual meeting of the *Linguistic Society of America Conference*, Chicago, January 2000.
- Kahnemuyipour, Arsalan. 2019. Word-internal modification: The case of Persian comparative marker. *Canadian Journal of Linguistics* 64. 138–145. <https://doi.org/10.1017/cnj.2018.27>
- Karvovskaya, Lena. 2013. 'Also' in Ishkashimi: Additive particle and sentence connector. In Felix Bildhauer & Mira Grubic (eds.), *Interdisciplinary Studies on Information Structure* 17, 75–97.
- König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective*. London: Routledge.
- Kornfilt, Jaklin, & John Whitman. 2011. Afterword: Nominalizations in syntactic theory. *Lingua* 121. 1297–1313. <https://doi.org/10.1016/j.lingua.2011.01.008>
- Lazard, Gilbert. 1992 [1957]. *A grammar of contemporary Persian*. Costa Mesa, CA: Mazda Publishers. (Translated from the original 1957 version entitled *Grammaire du persan contemporain* by Shirley A. Lyon.)
- Lazard, Gilbert. 1989. Le persan. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 263–293. Wiesbaden: Reichert.
- Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67. 475–509. <https://doi.org/10.1353/lan.1991.0009>
- O'Grady, William & John Archibald. 2016. *Contemporary linguistic analysis: An introduction*, 8th edn. Toronto: Pearson Canada, Inc.

- Oroji, Mohammad Reza & Amir Rezaei. 2013. Exploring *ke* as a focus particle in Persian from both form and function points of view. *Australian Journal of Linguistics* 33. 76–84.  
<https://doi.org/10.1080/07268602.2013.801108>
- Perry, John R. 2007. Persian morphology. In Alan S. Kaye (ed.), *Morphologies of Asia and Africa*, 975–1019. Winona Lake, IN: Eisenbrauns.
- Roberts, Ian & Anna Roussou. 1999. A Formal Approach to ‘Grammaticalization’. *Linguistics* 37. 1011–1041. <https://doi.org/10.1515/ling.37.6.1011>
- Schütze, Carson T. 2001. On Korean Case stacking: The varied functions of the particles *ka* and *lul*. *Linguistic Review* 18. 193–232. <https://doi.org/10.1515/tlir.2001.001>
- Song, Jae Jung. 1997. The so-called plural copy in Korean as a marker of distribution and focus. *Journal of Pragmatics* 27. 203–224. [https://doi.org/10.1016/S0378-2166\(96\)00042-2](https://doi.org/10.1016/S0378-2166(96)00042-2)
- Stilo, Donald. 2004. Coordination in three Western Iranian languages: Vafsi, Persian and Gilaki. In Martin Haspelmath (ed.), *Coordinating constructions* (Typological Studies in Language 58), 269–330. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.58.16sti>
- Taylor, John R. 2003. Polysemy’s paradoxes. *Language Sciences* 25. 637–655.  
[https://doi.org/10.1016/S0388-0001\(03\)00031-7](https://doi.org/10.1016/S0388-0001(03)00031-7)
- Traugott, Elizabeth Closs. 2007. Discussion article: Discourse markers, modal particles, and contrastive analysis, synchronic and diachronic. *Catalan Journal of Linguistics* 6. 139–157.  
<https://doi.org/10.5565/rev/catjl.128>
- Trousdale, Graeme. 2008. A constructional approach to lexicalization processes in the history of English: Evidence from possessive constructions. *Word Structure* 1(2). 156–177.  
<https://doi.org/10.3366/E1750124508000202>
- Van Bogaert, Julie. 2011. ‘I think’ and other complement-taking mental predicates: A case of and for constructional grammaticalization. *Linguistics* 49. 295–332.
- Windfuhr, Gernot L. 1979. *Persian grammar: History and state of its study*. The Hague: Mouton.  
<https://doi.org/10.1515/9783110800425>



# The pronoun-to-agreement cycle in Iranian Subjects do, objects don't

Geoffrey Haig

University of Bamberg

There is a broad consensus within linguistics that personal pronouns may undergo grammaticalization to yield person agreement morphology. Furthermore, it is widely assumed that similar processes apply to both subject and object pronouns. In this chapter I consider the fate of a phonologically identical set of clitic pronouns in Middle West Iranian languages, which were deployed in both subject and object indexing. The modern outcomes have been rather different; while erstwhile clitic subject pronouns have spawned subject agreement morphology in some languages, these clitic pronouns have not yielded obligatory object agreement in the category of person in any Iranian language. Neither traditional grammaticalization theory, nor recent formalizations of grammaticalization within Minimalism, offer a compelling explanation for this asymmetry. I suggest it reflects a fundamental difference in the informativity of subject as opposed to object indexing with respect to the category of person, as opposed to that of gender and number.

**Keywords:** grammaticalization, pronoun, agreement cycle, head-preference principle, Middle Persian, Old Persian, Central Kurdish, Vafsi, Sivand, Hawrami

## 1. Introduction

The view that long-term grammatical change is cyclic in nature was widespread among linguists of the late nineteenth and early twentieth centuries, for example Georg von der Gabelentz, Edward Sapir, and Otto Jespersen. More recently, the idea has resurfaced in derivational approaches to diachronic syntax, where for example Jespersen's Cycle in the emergence of negation markers has been re-framed in Minimalist terms (van Kemenade 2000; van Gelderen 2011b; among many others). Perhaps the most comprehensive study in this direction is van Gelderen (2011b), who analyses a number of different kinds of diachronic change in terms of

internally-motivated cycles. In this chapter, I focus on what van Gelderen (2011b) refers to as the ‘head-marking cycle’. This cycle begins with a pronoun, an element filling an argument position in syntax, which develops into an agreement marker, hence lacking a theta role, before finally eroding to zero. The cycle then begins afresh, with a new element emerging in the pronoun function. According to van Gelderen (2011a, 2011b), this cycle is attested both for subject pronouns, and object pronouns, and can be attributed to universally operative and internally motivated principles within the Minimalist framework.

The Iranian languages, with some 2,500 years of attested history, and dozens of surviving modern languages, provide a generous window for observing the kinds of long-term diachronic changes which cyclic approaches presuppose. In this chapter I briefly sketch what is known regarding the ‘agreement cycle’ in West Iranian languages, and evaluate the cyclic model of van Gelderen (2011b) against the Iranian evidence. I consider the development of subject pronoun to subject agreement marker, and for object pronouns to agreement marker respectively. The most striking finding is the almost complete absence of such a development for object pronouns, all the more surprising given the fact that the assumed preconditions for the grammaticalization of object agreement, namely cliticization of the relevant pronouns, has been available for millennia. But to the best of my knowledge, the early cliticization of object pronouns has not yielded object agreement anywhere in Iranian. For subject pronouns, on the other hand, uncontroversial cases of agreement markers developing from erstwhile clitic pronouns are attested, though only a small section of the assumed cycle is actually historically verified.

The asymmetry in the outcomes of subject and object pronoun grammaticalization in Iranian is not an isolated phenomenon, but reflects a widespread typological tendency. Nevertheless, much of the relevant literature continues to assume a unified grammaticalization pathway for subject and object pronouns. Thus Schiering (2005: 45) simply states that “cliticized subject pronouns can become agreement affixes cross-referencing the subject NP; cliticized object pronouns can become agreement affixes cross-referencing the object NP”, Siewierska (2004) suggests that the grammaticalization of pronouns towards agreement is “a continuous process on-going in all languages in all times”, without differentiating object and subject pronouns, while van Gelderen (2011b), assumes the existence of two cycles, a subject cycle and an object cycle, but provides no explanation for the evident cross-linguistic differences. In this chapter, I briefly outline the essence of van Gelderen’s (2011b) proposals, then present a summary overview of the relevant data from Iranian. Finally, I address the adequacy of the Minimalist cyclical approach to the grammaticalization of agreement. Given the scope of the issues involved, I can do little more than sketch the main arguments and present what I trust is a

reasonably representative cross-section of data. Some of the issues here have been dealt with from a cross-linguistic perspective in Haig (2018a, 2018b), to which the reader is referred for additional arguments and data.

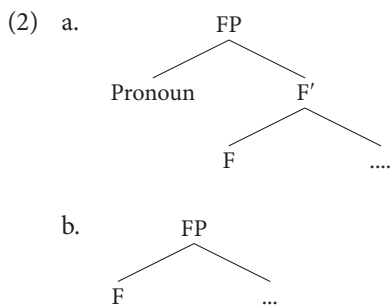
## 2. The agreement cycle according to van Gelderen (2011b)

The idea that verbal agreement markers arise from originally free pronouns was popularized by Givón (1976), and is rooted in the observation that in many languages, paradigms of agreement affixes often demonstrate close phonological similarities to the corresponding sets of personal pronouns (Siewierska 2004: 251–254; Haig 2018b). A natural explanation for these similarities is that the agreement affixes represent the grammaticalized remnants of erstwhile free pronouns. Indeed, this assumption is widely regarded as a given. As Siewierska (2004: 251) notes, “everyone acknowledges that person clitics and affixes typically evolve from independent person markers” [free pronouns, GH].

Van Gelderen (2011b) likewise assumes that pronouns are a common diachronic source for agreement morphology. In her framework, the development is seen as one of several cyclic processes in the creation of inflection, behind which quite abstract and very general principles can be identified. The theory is powerful in the sense that superficially distinct processes are considered as manifestations of a small number of very general principles. The most relevant principles in the present context are so-called Principles of Economy, which are operative in the resolution of “ambiguous structures” that arise in the derivation of syntax (van Gelderen 2011b: 13). Despite the name, Principles of Economy are not general cognitive principles geared to optimizing processing costs. Rather, they are principles specific to ‘I-Language’, rather than performance-based principles relevant to ‘E-Language’. In what follows, I will only consider two Principles of Economy, The Head Preference Principle (HPP) and Feature Economy (FE). The Head Preference Principle (HPP) is given in (1):

- (1) The Head Preference Principle (HPP)  
 Be a head, rather than a phrase (van Gelderen 2011b: 13)

More generally, “whenever possible, a word is seen as a head rather than a phrase” (van Gelderen 2011b: 13). The effects of the HPP can be schematically illustrated in (2), where FP stands for any functional category (here illustrated with a pronoun). When a functional element such as a pronoun or an adposition is merged, the HPP will yield an interpretation (2b), rather than (2a), if a speaker is exposed to evidence compatible with either (van Gelderen 2011b: 13).



(van Gelderen 2011b: 13 Example 16)

In terms of syntactic derivation, the HPP translates into a preference for head, rather than specifier position. With regard to the difficulties of distinguishing specifiers from heads, van Gelderen (2011b: 14) provides the following criteria: “Specifiers are full phrases and can be modified and coordinated, and they occur in certain positions; a coordinated or modified element is never a head, and head movement is usually recognizable.”

The HPP is relevant to a number of historical changes, for example demonstrative *that* > complementizer *that*, adverb > aspect marker, or pronoun > agreement. Of course changes of this kind are regularly cited in the grammaticalization literature, and accounted for in terms of a cline from ‘lexical to grammatical’, or ‘less grammatical to more grammatical’. The Minimalist account of van Gelderen (2011b) is an attempt to integrate these observations into a more formalized framework, and define more rigorously the somewhat vague notion of ‘more grammatical’.

The second kind of principle that is relevant for the subject agreement cycle concerns the nature of the features associated with the merging elements. Pronouns typically involve features that are both relevant for the semantic interpretation of an utterance, but also for the correct spell-out of associated inflectional morphology. In the version of Minimalism espoused in van Gelderen (2011b: 16–17), features are considered to be either ‘interpretable’ or ‘uninterpretable’:

Starting with Chomsky (1995), the features relevant for and accessible during the derivation are formal. Formal features can be interpretable (relevant to the semantic interface) or uninterpretable (only relevant to move elements to certain positions). Interpretable features are acquired before uninterpretable ones [reference omitted, GH<sup>1</sup>], but are later reinterpreted as uninterpretable, triggering the functional/grammatical system. The same happens in language change.

---

1. The reference deleted from this citation is to an internet publication, but the source provided in van Gelderen (2011b) is no longer identifiable online, hence I have removed it; it is not relevant to the arguments at hand.

Uninterpretable features are preferred because they provide the impetus for the derivation: “If you select two words from the lexicon with only interpretable features, they will not interact or merge.” (van Gelderen 2011b: 20)<sup>2</sup>

With regard to the pronoun-to-agreement shift, the changes can be schematically illustrated as follows, where ‘phi’ abbreviates the person values first, second and third person (1,2,3), ‘i’ abbreviates ‘interpretable’, and ‘u’ is ‘uninterpretable’:

- (3) emphatic full pronoun head pronoun agreement  
 [i-phi] > [i-phi] > [u-1/2], [i-3] > [u-phi]

The cline sketched in (3) is driven by a Feature Economy Principle, formulated in van Gelderen (2011b: 17) as “Minimize the semantic and interpretable features in the derivation”. Van Gelderen (2011b) also distinguishes between a feature ‘first/second person’ and ‘third person’, a move motivated by the fact that pronouns with these features grammaticalize at different rates towards agreement; in general, first and second person pronouns spearhead the development, becoming uninterpretable earlier, while third person pronouns apparently lag behind. On her view, the pronouns of the first and second person entail “pure phi-features (person and number)” (van Gelderen 2011b: 74). Third person pronouns on the other hand, encode additional features, though the nature and number of these features is a matter of typological variation. Typically they involve gender, and deixis; the latter would be particularly true of languages lacking dedicated third person pronouns, instead relying on forms identical to distal demonstratives. The forms with the simplest feature specification are therefore first and second person forms, and these are the forms which are thus more likely to shift their features from interpretable to uninterpretable.

Taken together, the Head Preference Principle and Feature Economy conspire to nudge free pronouns, as phrase-projecting carriers of interpretable features, to become exponents of non-projecting heads with uninterpretable, but syntactically relevant, features: agreement morphology. Of course in order to complete the cycle, the agreement morphology must further develop to zero, before being replaced by innovated material realizing the relevant features, thus yielding a complete cycle. The shift from agreement head to zero may apparently be triggered by the ‘stacking up’ of additional material in the same slot, as additional functional heads accrue in the same position leading to opaquely fused morphology, and ultimately complete loss of the original material (van Gelderen 2011b: 19–21). However, this stage of the cycle (the loss of agreement) will not be considered further here.

---

2. The Feature Economy Principle outlined here is actually considered an offshoot of ‘Late Merge’, but I have omitted the relevant argumentation here, see van Gelderen (2011b: 14–17) for details.



Full coverage of the technical details of van Gelderen's proposals would extend beyond what can reasonably be accommodated here.<sup>3</sup> The crucial point is that the explanations are formulated in terms of purely syntactic and very general principles, which should in principle be universally operative.

## 2.1 The agreement cycle with subject pronouns

Strictly speaking, the starting point of the agreement cycle in (3) are 'emphatic pronouns', but I will assume here 'full pronouns' for reasons of brevity. Full pronouns exhibit certain properties, which are exemplified with pronouns from Hindi in (4). They can be extended with an additional noun, as in (4a)–(4b), or carry a special focus particle (4c), they inflect for case in the same manner as other DPs, show similar positional distribution to DPs, and can also be coordinated. These are typical diagnostics for the 'nominal' nature of such pronouns: "In terms of features, the pronouns and DPs are full phrases at this stage and carry the traditional Case and phi features." (van Gelderen 2011b: 47)

- (4) *mē* "I", *tum* "thou", *wo* "she/he", *ham* "we", *aap* "you", *wo* "they"  
 a. *ham log* "we people"  
 b. *aap log* "you people"  
 c. *mē hii* [1SG-FOC] "I" [Hindi, van Gelderen 2011b: 45]

The transition to the second stage, so-called 'head pronouns', is not abrupt; instead we find various intermediate phases. The so-called 'subject pronouns' of English (*I*, *you*, *he*, *she*, etc.) differ, for example, from the so-called 'object pronouns' (*me*, *you*, *him*, *her*, etc.) with respect to some of the relevant parameters. The subject pronouns are generally unstressed, have less syntactic freedom (they are restricted to a position immediately preceding a finite verb, separable from it only by a small set of adverbs), not available in isolation or in focus constructions such as *as for X*, ..., and are dispreferred in coordination (a combination such as *?they and we*, for example would be avoided in my dialect of spoken English). On van Gelderen's (2011b) approach, this is evidence of an initial move down the subject agreement cycle, from full to head pronoun. More advanced developments are found in colloquial French, where the weak series of pronouns *je* "I", *tu* "you", *lui/elle* "he/she", etc. frequently double an overt subject NP, as in (5) and (6):

---

3. I also ignore the discussion of so-called polysynthetic languages, for which somewhat different mechanisms are required (van Gelderen (2011b: 43–44); see Corbett (2006: 100–113) for critical discussion of 'pronominal affixes', which is relevant to the notion of polysynthesis.)

- (5) *une omelette elle est comme ça*  
 an omelet she is like this  
 “An omelet is like this.” [Spoken Swiss French, Fonseca-Greber 2000: 335,  
 cited in van Gelderen 2011b: 52]
- (6) *Moi je suis un blogueur*  
 Me I am a blogger  
 “I am a blogger.” [colloquial French, van Gelderen 2011b: 53]

There is evidently good reason to consider the weak pronoun set of French *je, tu, lui/elle*, etc. to be functionally distinct from the free pronouns of, e.g. Hindi discussed in (4). In van Gelderen’s terms, they can be considered a further step on the cycle towards agreement.

The final stage of fully obligatory agreement marking is illustrated by languages such as German or Latin, exhibiting obligatory agreement with subjects, regardless of any pragmatic considerations. It is worth pointing out that precisely this kind of canonical agreement (Corbett 2006) often cannot be directly traced to a pronominal predecessor. Van Gelderen (2011b) does not actually provide a convincing example of the entire cycle, but instead takes the structures from (often unrelated) languages as representatives of the various stages of the assumed cycle. The assumption seems to be that the weak pronouns of French illustrated in (5)–(6) will somehow eventually morph into obligatory agreement affixes, given sufficient time, but clear evidence of such a process is hard to find, as Siewierska (1999) had already noted. As it turns out, evidence for the final stages of this process can be found in Iranian (see § 3.1).

## 2.2 The agreement cycle with object pronouns

Direct evidence for the object agreement cycle is hard to come by. Van Gelderen (2011b) illustrates it with the following fictitious example:

- (7) a. I saw yesterday her (and him)  
 b. I saw’r yesterday (\*and him)  
 c. I saw (’r) HER. [Fictitious English, van Gelderen 2011b: 88]

In (7a), the pronoun is syntactically and prosodically independent (separable from the verb by an adverb), bears a theta role, and can be coordinated. In (7b) it has lost positional freedom and prosodic independence, and also the ability to coordinate. In (7c) we observe the possibility of doubling the attached pronoun through an additional ‘emphatic pronoun’. At this stage the attached pronoun may erode to zero, “and the cycle can start over again.” (van Gelderen 2011b: 88). The stages of these developments are sketched in the form of “a possible cline” in (8):

- (8) phrase > head > agreement > zero  
 [i-phi] [i-phi] [u-phi] [u-Case] (van Gelderen 2011b: 88)

Van Gelderen (2011b: 90) points to the considerable cross-linguistic diversity in object agreement systems: “With respect to object agreement, there is enormous diversity as to what starts the cycle. Animate and definite object pronouns of all persons are reanalyzed as object agreement but there is no obvious pattern.” Similarly, the initial structural configuration for objects is less clear-cut, because current conceptualizations of VP structure yield somewhat different analyses (cf. the alternative options in van Gelderen 2011b: 89–90). But in essence, both the HPP and Feature Economy are considered active in driving the developments, just as they are with subject pronouns. A problem nevertheless arises with regard to Feature Economy, because object agreement is often sensitive to animacy, and definiteness, features that are related to person, but not identical to it. Van Gelderen (2011b: 90) assumes that definiteness is related to the presence of uninterpretable [ASP] (aspect) features on the verbal head governing the object, though I find the connection somewhat tenuous. But apart from the apparent difficulties in identifying the favoured starting configuration for the cycle, van Gelderen (2011b) assumes that the object agreement cycle can be motivated in a manner that parallels that of the subject agreement cycle, and there is thus no expectation that the outcomes of the two cycles will be any different.

### 3. Clitic pronouns and agreement in Iranian

Subject agreement via affixes on the verb is present in most, perhaps all, modern Iranian languages, though it may be absent for past transitive clauses. For ease of exposition, we may take modern standard Persian, where the paradigm of subject agreement suffixes is provided for the present indicative in Table 1.

**Table 1.** Subject agreement suffixes in Persian (present indicative of *xordan* “eat”)

	Singular	Plural
1	<i>mi-xor-am</i>	<i>mi-xor-im</i>
2	<i>mi-xor-i</i>	<i>mi-xor-id</i>
3	<i>mi-xor-ad</i>	<i>mi-xor-and</i>

The markers themselves are suffixal, rather than clitics: they are restricted to a specific slot (immediately following the verb stem), they are obligatory in the sense that they are required by a particular syntactic configuration, irrespective of the presence or absence of a full NP subject in the clause, and according to Kahnemuyipour

(2003: 374–375), are ‘cohering’ suffixes, i.e. part of the phonological word. In the sense of van Gelderen (2011b), they are syntactic heads, associated with uninterpretable phi-features.

The origins of these suffixes are obscure. Persian has exhibited some form of agreement suffixes in comparable environments for as long as we have attested records. Thus if they are the endpoint of a grammaticalization process that began with a free pronoun, the earlier phases of this development lie beyond the bounds of what can realistically be reconstructed.

Along with the suffixes of the type illustrated in Table 1, Persian and the majority of other Western Iranian languages exhibit a second set of prosodically dependent person and number marking morphemes, often referred to as clitic pronouns. Although the paradigms are not fully identical, they can reasonably be considered cognate with the pronominal clitics attested in Middle West Iranian language such as Parthian, Middle Persian and Bactrian (Jügel 2015). The Middle West Iranian clitics are provided, together with a selection of contemporary West Iranian languages in Table 2.<sup>4</sup>

Table 2. Clitic pronouns in Western Iranian languages

	Appr. 2,000 years BP	Contemporary West Iranian languages				
	Middle West Iranian	Persian	Vafsi	Hawrami	Sivand	C. Kurd. Sanandaj
1SG	= <i>m</i>	= <i>am</i>	= <i>om</i>	=(1) <i>m</i>	= <i>em</i>	= <i>im</i>
2SG	= <i>t</i>	= <i>at</i>	= <i>i</i>	=(1) <i>t</i>	= <i>et</i>	= <i>it</i> / = <i>o</i>
3SG	= <i>š</i>	= <i>aš</i>	= <i>es</i>	=(1) <i>š</i>	= <i>eš</i>	= <i>ī</i>
1PL	= <i>mān</i>	= <i>mān</i>	= <i>oan</i>	= <i>mā</i>	= <i>emā</i>	= <i>mān</i>
2PL	= <i>tān</i>	= <i>tān</i>	= <i>ian</i>	= <i>tā</i>	= <i>etā</i>	= <i>tān</i>
3PL	= <i>šān</i>	= <i>šān</i>	= <i>esan</i>	= <i>šā</i>	= <i>ešā</i>	= <i>yān</i>

This paradigm of clitic pronouns has proved remarkably robust, surviving across at least 2,000 years in recognizable form in the majority of West Iranian languages, though lost in Zazaki, Northern Kurdish, Gilaki and Mazandarani. Note that these clitic pronouns are not simply phonologically reduced forms of today’s full pronouns. Rather, they are the reflexes of a distinct set of clitic non-nominative pronouns, of which the corresponding full pronouns have disappeared. From a

4. Sources for the languages other than Persian: Middle West Iranian: Jügel (2015: 222); see also Korn (2009) for historical details on this paradigm; Vafsi: Stilo (2018: 695, Table 5E); Hawrami: MacKenzie (1966: 25); Sivand dialect: Lecoq (1979: 40); Central Kurdish of Sanandaj: Öpengin & Mohammadirad (to appear). Apparent differences in the qualities of the vowels are in part due to differences in the transcription practices of the sources; they are irrelevant for the present purposes.

synchronic perspective, they are not relatable to the full forms of the pronouns via predictable phonological rule in any of the languages listed.

For reasons outside the purview of this chapter, in the Old Iranian period, the clitic pronouns in Table 2 came to be used as **subject** pronouns with past transitive verbs. I follow a long tradition in referring to this function as ‘A’, meaning ‘transitive subject’, but it should be borne in mind that the only transitive subjects that occurred with this kind of pronoun were those associated with verbs built on the old participial stem, generally referred to as the ‘past stem’.<sup>5</sup> In the next section I briefly sketch the workings of these subject clitic pronouns, while in § 3.2, I discuss the same set of pronouns in object function.

### 3.1 Clitic pronouns indexing subjects (A)

In Old Iranian, and well into Middle Iranian, the subject clitic pronouns were in complementary distribution with a co-referent NP subject. Example (9) from Middle Iranian illustrates a clitic pronoun A, while (10) has a NP in the A role, and no clitic pronoun:

- (9) *čē=t ātaxš ī man pus ōzad*  
 because=2SG:A fire of my son extinguish.PST.3SG  
 “because you extinguished the fire of my son ...”

[Middle Persian, Haig 2008: 124]

- (10) *pas ošbām oy az pidar bōxt ...*  
 then ošbām:A 3SG from father rescue.PST.3SG  
 “then Ošbām rescued her from (her) father ...”

[Zoroastrian Middle Persian, Jügel 2015: 410, glosses added]

Jügel (2015: 400) notes the general lack of clitic doubling in Middle Iranian,<sup>6</sup> underscoring the pronominal nature of the clitics at this stage. Another important indicator of their pronominal nature is that they could be omitted in contexts where the identity of the subject is pragmatically recoverable, for example in same-subject clause chaining. Example (11) has an overt clitic pronoun for the A of the first clause, and zero for the co-referential A of the subsequent clause:

---

5. In fact they also serve as ‘subject’ agreement for a number of other constructions across the West Iranian languages, including predicates of experience, desire, possession, and physical and mental states such as hunger. In these functions, the use of the clitic pronoun as a subject index is independent of tense, see Haig & Adibifar (2019).

6. Jügel (2015: 396–399) notes a small number of Middle Persian examples where the A-clitic is doubled by an overt A in the clause (94 attested in a corpus of 6,815 clauses). Some may be attributed to scribal errors or other problems of interpretation and transmission.

- (11) a. *u=š ardawān ōzad ...*  
 and=3SG:A Ardawān kill.PST.3SG
- b. *ud duxt ī ardawān pad zanih kard*  
 and daughter of Ardawān to wife make.PST.3SG  
 ‘And he<sub>i</sub> killed Ardawān ... and (he<sub>i</sub>) took his daughter as wife’  
 [Zoroastrian Middle Persian, Jügel 2015: 411, glosses added]

The available evidence thus supports a pronominal interpretation of the clitic pronouns, because (i) they cannot co-occur with a co-referent NP in the same clause, and (ii) they may be omitted in precisely those environments that free pronouns would likewise generally be omitted (e.g. coreferential deletion in coordinate clauses).

However, there are criteria for distinguishing among different kinds of pronoun, rather than assuming a general binary split between an agreement marker on one hand, and a pronoun on the other (cf. Jügel & Samvelian, this volume, for the latter view). Van Gelderen (2011b) recognizes a distinction between free pronouns, and ‘head pronouns’. The former have the same word order freedom as lexical NP’s, are stressable, focusable, and can be modified and coordinated (cf. discussion in connection with (4) above). Head pronouns, on the other hand, lack at least some of these features. The clitic A-pronouns of Middle Iranian would most likely qualify as ‘head pronouns’: their position is fixed through the second-position principle governing clitic placement in Middle Iranian, and it seems unlikely that they were stress-bearing, or capable of expressing contrastive focus.

There is a further criterion for distinguishing between free pronouns and head pronouns, discussed in Haig (2018a: 67). As mentioned, free pronouns are characteristically omitted under conditions of pragmatic identifiability of the referent, and this can be considered a general feature of pronouns, though famously subject to cross-linguistic variability (see Torres Cacoullós & Travis, 2019; Haig & Adibifar 2019). Thus full pronouns are typically characterized by a pragmatically-determined alternation with zero. The clitic subject pronouns of Middle Iranian could also be omitted, for example in the second conjunct of same-subject clause sequences (cf. (11b) above). But in fact, several Middle Iranian examples illustrate clitic pronouns in contexts where pronouns would not normally be expected, for example the following:

- (12) *ēk, ke=š man brēhēnīd*  
 one, that=3SG.A 1SG create.PST.3SG  
 ‘one which created me’ [*lit.* ‘one that he created me’],  
 [Zoroastrian Middle Persian, Jügel 2015: 378, glosses added]

The subject pronoun =š attaches to the relativizer *ke*, although resumptive pronouns are generally not required in Iranian subject relativization. Jügel & Samvelian (this volume, § 2.1) also note the propensity for clitic pronouns to occur in same-subject

sequences of main and embedded clause in their Middle Persian corpus, again precisely an environment where zero would be the expected option. Jügel & Samvelian (this volume, Figure 5) provide figures from the analysis of a single Middle Persian text. The numbers of zero subjects in past transitive clauses is significantly lower than in present transitive clauses (44% versus 72%). A Fisher's exact test of this difference yields a very significant value of 0.0001.<sup>7</sup> What I would provisionally conclude from these findings is that the past transitive clauses avoid zero expression of subjects to a greater degree than the present transitive clauses.

The overall figures for the Middle Persian corpus investigated in Jügel (2015: 326, Table 5.4) indicate that around 44% of all past transitive clauses contained a clitic pronoun exponent of the subject ( $N = 6,815$ ). Comparing this figure with the percentage of overt pronouns in transitive clauses of other languages which allow referential null subjects is revealing: In contemporary spoken Persian (Adibifar 2016), overt A pronouns occur in 8% ( $N = 603$ ) of the transitive clauses, in Cypriot Greek (Hadjidas & Vollmer 2015) we find just 4% ( $N = 494$ ), and in Northern Kurdish (Haig & Thiele 2015) 29% ( $N = 422$ ).<sup>8</sup> For these languages, and indeed most others that allow null referential subjects, the favoured form of expression for transitive subjects is zero, not pronominal. The Middle Persian figure of 44% (a conservative estimate) is thus significant, and suggests that these so-called 'pronouns' were of a qualitatively different kind to the free pronouns.

Jügel & Samvelian (this volume) also note the difference, and assume that it is due to the lack of subject agreement morphology on past transitive verbs. This suggests that the triggering factor for the grammaticalization of clitic pronoun subjects towards agreement markers was essentially structural: the loss of an old paradigm of suffixal agreement morphology is compensated by recruiting a new paradigm from the available clitic pronouns. This scenario is in line with Fuß' (2005) claims regarding the motivation for the emergence of subject agreement as compensation for defective agreement paradigms. In principle I find this plausible, and the

---

7. My calculations are based on a comparison of non-zero realizations (NP, pronouns) versus zero-realizations of A-arguments in two conditions, past versus present tense (based on the figures in Figure 5, Jügel & Samvelian, this volume). Interestingly, with intransitive subjects there is an inverse effect of tense, with zero realizations being more frequent in the past than in the present. This appears to be linked to the more narrative nature of the past-tense sections of the text (cf. Jügel & Samvelian, this volume), which would favour topic continuity over longer stretches, hence zero expression. If this is the case, then it further heightens the significance of the reduced levels of zero realizations for transitive subjects in past tenses.

8. The comparatively high figure for Northern Kurdish is probably related to the fact that in the Northern Kurdish corpus, many of the verbs are past tense transitives, which lack overt agreement morphology.

frequency data from Middle Iranian provide empirical support for such a view. To what extent additional explanations in terms of re-analysis of ‘hanging topic’ constructions are required remains an open question (see Schnell 2018, among many others, for critique of the ‘dislocated topic’ approach to the emergence of agreement, and Jügel & Samvelian, this volume, for an attempt to justify it for Iranian).

The system of indexing the A through a pronominal clitic has disappeared in some contemporary Iranian languages, notably Persian, but in others it has survived remarkably well. However, in some languages the nature of the clitic pronoun has changed. In Central Kurdish, the pronominal clitic has become fully obligatory: “every single past transitive construction requires an A-past clitic”, regardless of the presence or absence of an overt A constituent in the same clause (Haig 2008: 288). Along with a functional shift, the clitics have changed their position, from the clause-second position of Middle Iranian to a VP-based placement (cf. Haig’s (2008: 336) ‘rightward drift’ of clitic placement in Iranian). This is illustrated with the following examples from the Mukri dialect of Central Kurdish: (13) shows the co-occurrence of a pronominal clitic with a definite NP subject, (14) an indefinite, non-specific subject, and (15) a pronominal subject.

- (13) *qerewol-ān kut=yān*  
*guard-PL say.PST=3PL*  
 “The guards said ...”  
 [Öpengin 2013: 307, cited in Öpengin & Mohammadirad, to appear]
- (14) *hič kes řā=y-ne-de-girt-im*  
*no person PVB=3SG-NEG-IPVF-keep.PST-1SG*  
 “Nobody would let me in (their house).”  
 [Öpengin 2013: 51, cited in Öpengin & Mohammadirad, to appear]
- (15) *min ne=m-dît*  
*1SG NEG=1SG-see.PST*  
 “I did not see him.”  
 [Öpengin 2013, cited in Öpengin & Mohammadirad, to appear]

There is broad consensus in the relevant literature that the pronominal clitics in Central Kurdish are exponents of an agreement relation (see Samvelian 2007; Haig 2008, 2018b; Öpengin, 2019; and Öpengin & Mohammadirad, forthcoming; see Dabir-Moghaddam 2008 for examples from other West Iranian languages).

In a number of other West Iranian languages, however, the clitic remains in complementary distribution with a coreferent NP subject. This appears to be the situation in the Surčî dialect of Northern Kurdish, spoken in Iraqi Kurdistan:



- (16) *min la\_bo xo rēnjbar-ak girt, hinār=im=a jot,*  
 1SG.OBL for self labourer-INDEF take.PST.3SG send=1SG=DRCT plough  
*šiwān-ak=iš=im girt*  
 shepherd-INDEF=ADD=1SG take.PST.3SG

“I hired a labourer, I sent (him) to the plough, (then) I hired also a shepherd.”  
 [MacKenzie 1962: 228, cited in Öpengin & Mohammadirad, to appear]

In this dialect then, unlike the Central Kurdish outlined in (13)–(15), a subject clitic is not obligatory. In fact there are also clauses in MacKenzie’s (1961) data that contain neither a subject clitic, nor an overt subject NP, indicating that the clitic pronouns are still omissible under pragmatically felicitous conditions.<sup>9</sup>

Finally, we can point to those languages where the clitic pronouns are used for subject indexing, but have lost all positional freedom and occur exclusively on the verb stem itself, thus resembling more closely an affix. This is found with third person subjects in the Kakevendi and Aleshtar dialects of Lak, where the subject clitic only occurs on the verb, regardless of the availability of other potential hosts in the clause:

- (17) *tamām māhī-la hwārd=ē*  
 all fish-PL eat.PST=3SG  
 “He ate all the fish.”

[Lak of Kakevendi, Öpengin & Mohammadirad, to appear]

In the Central Plateau dialect of Semnān, the subject clitics (with past transitive verbs only) have entirely lost their syntactic mobility, and are now restricted to occurring on the verb stem (Haig 2018a).

In sum, across Western Iranian we witness the presence of clitic pronouns indexing past transitive subjects. These pronouns were originally special clitics, rather than free pronouns. They were syntactically constrained in their placement possibilities, and thus presumably lacked the ability of free pronouns to express contrastive focus, or to be coordinated. They also differed from free pronouns in having a higher overall frequency of occurrence, though this requires further investigation. However, they remained pronominal in the sense that they were in complementary distribution with co-referent NP subjects. Pronominal clitics exhibiting very similar properties can still be observed in the Surči dialect of Northern Kurdish, and in at least some dialects of Hawrami, though there are complications here involving word-order variation, and case-marking. For Central Kurdish, on the other hand, and perhaps for Semnān dialect, the clitic pronouns are now fully obligatory agreement markers.

9. See for example “... *sē zēr ta kūna karīrā kird*. ... (he) put three (pieces of) gold up the donkey’s backside” (MacKenzie 1962: 232).

The development could thus be interpreted as traversing a sub-section of van Gelderen's subject agreement cycle, namely that of head pronoun to agreement marker, discussed in (3) above and repeated here for convenience:

- (3) emphatic full pronoun head pronoun agreement  
 [i-phi] > [i-phi] > [u-1/2], [i-3] > [u-phi]

As mentioned, the status of van Gelderen's 'head pronoun' is somewhat obscure. Likewise, as yet we lack evidence for the assumed split of first and second person versus third person pronouns. But the clitic pronouns that we encounter in Middle Iranian (9)–(12) are not emphatic pronouns, and arguably distinct from full pronouns. Thus over 2,000 years, we find in some, but not all, languages, evidence for the assumed development from pronoun towards agreement marker, albeit only involving the final stages of the cline, and only attested in some of the languages that have the relevant pronouns. The presumed initial stages, i.e. the development of full pronouns to clitic pronouns, lies beyond the bounds of the historical records. Thus at least 2,000 years were required for just the final section of the assumed developmental cline to unfold.

### 3.2 Clitic pronouns indexing objects

The use of clitic pronouns for objects is a feature that characterizes Iranian back to its earliest attestation, so we can assume their presence in Iranian languages for at least 2,500 years and probably longer. As such, they are of greater antiquity than the clitic pronouns for subjects (previous section), which only emerged in the wake of the shift to ergative alignment, and only in past tenses (Haig 2008; Jügel 2015). But despite the fact that clitic object pronouns have been around longer, I claim that they have not moved significantly closer towards agreement than their earliest attested forerunners.

In Old Iranian, there was still a dedicated paradigm of accusative clitic pronouns, which later syncretized with the other non-nominative clitic pronouns to yield the paradigm provided in Table 2 above. Examples of Middle Iranian clitic pronouns in object function are given below (from Haig 2008: 115):

- (18) *čīd=mān pāyēd*  
 always=1PL protect.PRS.3SG  
 "(It) always protects **us**"
- (19) ... *u=š hamēw bōžēnd*  
 ... and=3SG always save.PRS.3PL  
 "(the Gods) always save **him**"

As can be seen, the clause-second placement principles also apply to the object clitics of Middle Iranian. In many contemporary Iranian languages, object clitics continuing the Middle Iranian ones just mentioned are found, though their placement principles have shifted. The following examples illustrate the position of the object clitics in the Mukri dialect of Central Kurdish (Northwest Iranian, West Iran, Öpengin 2016). The clitic attaches to (approximately) the first stress-bearing constituent<sup>10</sup> of the VP, which could be a negation or a modal prefix as in (20a)–(20b):

- (20) *kut=i*                    “*segbāb bo de=m=guž-i?*”  
 say.PST=3SG.A dog.son why IND=1SG=kill.PRS-2SG  
 “He said: ‘Son of a dog, why are you killing me?’”  
*kut=im*                    “*bāb=im nā=t=guž-im*”  
 say.PST=3SG.A brother=POSS1SG NEG=2SG=kill.PRS-1SG  
 “I said: ‘O brother, I am not killing you’”            [Öpengin, 2016, ŽB 183–184]

The object clitics in (20a)–(20b) appear to be morphologically incorporated into the respective predicates, and would thus superficially at least seem to be highly grammaticalized. However, the degree of prosodic and morphological integration into the predicate is not matched by functional status as agreement: they are not obligatory, and do not double an overt NP object.

Although cognate sets of object clitic pronouns are attested in numerous Western Iranian languages, the descriptions I am aware of show that the clitic object pronoun is in complementary distribution with an overt NP object, regardless of the degree to which the clitic pronoun is phonologically and morphologically integrated into its host. The best-known counter-examples to this trend come from colloquial spoken Persian, where sporadic instances of clitic doubling can be found. Van Gelderen (2011b: 96) cites examples from Lazard (2006 [1957]), which apparently illustrate that object clitic pronouns in Persian are moving towards agreement (cited from the reprint (2006), and rendered in the colloquial style of transcription):

- (21) (*to xodet miduni ...* “you yourself know...”)  
*ke man to=ro duss=et dār-am*  
 that I 2SG=ACC loving=2SG have.PRS-1SG  
 “... that I love you”            [colloquial Persian, Lazard 2006 [1957]: 100, 176]

The other examples cited involve third person objects, such as the following:

10. This is an over-simplification, as the indicative prefix in (16a) is not in fact stressed; see Öpengin (2019) for a detailed discussion of the clitic placement in Mukri Kurdish.

- (22) (*umadan mixan ... “they came wanting...”*)  
*baba-jun=o be-gir-an=eš*  
 father-dear=ACC SUBJ-arrest.PRS-3PL=3SG  
 “... to arrest dear father” [colloquial Persian, Lazard 2006 [1957]: 176]

Both these examples are actually taken from written works of fiction (Čubak). Although the author from which these examples are taken is known for evoking the colloquial spoken language in his writing, it is at best an attempt to reflect the informal spoken language. A more reliable source for spoken Persian would be original utterances, such as (23):

- (23) *yek pesar-i āmad bā dočarxe ke yeki az*  
 one boy-INDEF come.PST.3SG with bike COMPL one of  
*zambil-hā=rā gozast=aš ruye dočarxe=aš*  
 basket-PL=ACC put.PST.3SG=3SG onto bike=3SG.POSS  
 “a boy came with a bike, then put one of those baskets onto his bike”  
 [Adibifar 2016, G2\_f\_7, 007, cited in Haig 2018b]

Van Gelderen (2011b: 96) considers examples such as (21)–(23) as evidence for the “reanalysis of the verbal object clitic as third person agreement.” However, it is not the case that all definite third person objects are accompanied by the corresponding clitic pronoun. The figures from the corpus of spoken Persian in Adibifar (2016) indicate that of the total number of 628 direct objects, just 46 are expressed through clitic object pronouns. Among those 46 cases of clitic object pronouns, a sole example, (23) above, involves clitic doubling (Haig 2018b). Thus more than 90% of direct objects are not indexed by a clitic pronoun at all, and doubling of the clitic pronoun with an overt NP is very unusual, at least in this corpus.

Although some examples are acceptable to native speakers, the frequency of direct objects that are accompanied by clitic doubling in actual usage is low. Furthermore, there are quite strict constraints on doubling; Rasekh (2014) notes that doubling the object clitic is not possible with objects that are indefinite, or in focus. My impression is that it is most natural with third person objects, less so with second person objects, and almost unacceptable with first person objects,<sup>11</sup> though this requires much more detailed investigation.

Van Gelderen (2011b), however, interprets the isolated instances of clitic doubling cited in the literature as evidence for an ongoing shift towards object agreement in Persian. There are several problems with this claim. First, there is actually

11. I am very grateful to Mohammad Rasekh-Mahand for sharing his intuitions as a native speaker and linguist on these constructions. Of course he bears no responsibility for how I have interpreted them.

no evidence that object doubling in contemporary colloquial Persian is an innovation. We simply do not know very much about the colloquial spoken Persian of earlier centuries; it was not written down or recorded. It is quite possible that object doubling has been available as a marginal stylistic device, linked to some specific pragmatic contexts, for centuries, perhaps millennia. Second, even if it should be a comparatively recent innovation, it will not necessarily inevitably proceed towards obligatory agreement. It has been shown that cross-linguistically, some kind of pragmatically conditioned object indexing, as opposed to obligatory agreement, is actually the norm (Siewierska 1999; Haig 2018b). As mentioned, clitic object pronouns are widespread across West Iranian languages. But to my knowledge, not a single language has developed obligatory object agreement based on these pronominal forms.

Note finally that object agreement *per se* is not ruled out in Iranian. A number of languages have obligatory agreement with objects in past transitive constructions, including Pashto, Northern Kurdish, or Zazaki, illustrated in (24):

- (24) *mi nā keynekī to-rē ārdā*  
 1SG.OBL this girl (FEM.) 2SG-for bring.PST.FEM.SG  
 “I brought **this girl** for you” [Zazaki, Paul 1998: 129, glosses added]

The agreement morpheme on the verb in (24) is not etymologically related to the third person singular clitic pronouns of Table 2, or their cognates in the other languages we have been discussing. Furthermore, object agreement is primarily in the features of gender and number, rather than person. This is typical for object agreement in Iranian: where it is found, it is not etymologically related to the clitic pronouns of Table 2, and most consistently indexes the features of number, and gender, rather than person (Haig 2017, 2018b). Thus object agreement in Iranian is attested, but it has not arisen via the grammaticalization of clitic pronouns in the same manner as subject agreement has. Given the antiquity of clitic object pronouns, and their wide distribution throughout Iranian languages, the lack of object agreement derived from object pronouns is highly conspicuous, and indicative of deep differences between object and subject agreement, though obscured by the superficial similarity in form between clitic object and clitic subject pronouns.

#### 4. Conclusions

Following changes in Iranian morphosyntax between the Old and the Middle Iranian period (perhaps 2,000–2,500 years ago), a paradigm of clitic pronouns (Table 2) came to be used to index past transitive subjects. In Middle Iranian, these subject clitic pronouns were in complementary distribution with free NP subjects;

this kind of system is still attested in some West Iranian languages to this day. In others, the subject clitic pronouns have become fully obligatory agreement markers, illustrated for Central Kurdish in (13)–(15). An identical paradigm of clitic pronouns has been used to index direct objects for even longer, and clitic object pronouns remain widespread across West Iranian to this day. But nowhere have they reached an agreement stage, despite their lack of prosodic independence and in some cases, morphological integration into the governing predicate.<sup>12</sup>

The history of Iranian provides thus some support for the Minimalist account of the grammaticalization of subjects, but only the final stages in the cycle (3) are actually attested; the assumed initial stages are beyond the realms of historical attestation. We can assume that a full cycle – from emphatic pronoun to subject agreement – would involve a time span in the realm of several millennia; this would explain why an unbroken chain of attestation covering all stages of the cycle is unlikely to be forthcoming for any language. For object pronouns, however, despite the presence of a seemingly optimal configuration for the start of the cycle, there is little evidence of further developments in the predicted direction.

Why should the outcomes of the two processes have turned out differently, despite the phonological identity of the input material? I have suggested elsewhere (Haig 2018b), that there is a strong typological tendency for object agreement to be conditioned, e.g. through definiteness, topicality, or animacy of the object, rather than be fully obligatory. Subject agreement, on the other hand, tends to be across-the-board obligatory. And where obligatory object agreement is attested, it is most frequently in number and gender, rather than person. Baker (2011) provides a partial explanation for the latter tendency from a synchronic perspective, but in this chapter we are centrally concerned with the diachronic mechanisms by which pronouns (may) become agreement markers (not all agreement markers originate from pronouns). The Minimalist account of van Gelderen (2011b) assumes general principles such as the Head Preference Principle, and Feature Economy as the driving forces behind the grammaticalization of pronouns. But neither would predict any differences between subject and object grammaticalization. An alternative

---

12. An anonymous reviewer points out that object pronouns do appear to be the source of object agreement markers in a number of languages, including Basque and Georgian. Two points need to be emphasized in this connection: First, conditioned object indexing (i.e. objects are only additionally indexed on the verb when they are topical, first and second person, human etc.) is widespread, and many scholars refer to it as agreement; this may also be true of Basque and Georgian, and would need to be established. My claim is simply that obligatory (unconditioned) agreement that is historically based on object pronouns is significantly less frequent than it is for subject agreement in Iranian, and this appears to echo a typologically more widespread tendency. Second, as noted in Haig (2018b), object agreement tends to be more likely in languages with ergative or at least non-accusative alignments, both of which would apply to Basque and Georgian.

usage-based explanation is set out in Haig (2018b), which points to the differing informativity of subject and object indexing with regard to the feature of person: the person value of a subject index is not readily predictable, while that of an object index (cross-linguistically, it appears that upwards of 90% of objects in discourse are actually third person; see Haig 2018b: 810–812 for details). Whether this can be confirmed remains an open question, but any account of the grammaticalization of pronouns towards agreement needs to account for the fundamental differences between subject and object pronouns in this regard.

Another important point to emerge from the Iranian data is that cliticization by itself is not necessarily the start of the slippery slope towards grammaticalization into inflectional morphology. Clitics can remain just that for millennia; there is nothing inevitable in the assumed clines for the grammaticalization of pronouns (see Schiering 2005 on the independence of phonological attrition and functional grammaticalization). This appears to be particularly true of object pronouns, which are frequently prosodically weak and attach to a verbal head, even in English. But they may evidently plateau at that stage for a very long time. The title of Siewierska (1999) sums this up poignantly: object pronouns just don't 'make it' to the assumed end of the grammaticalization cline. I see no compelling grounds for assuming that sporadic cases of object clitic doubling in Persian represent the first stage towards wholesale object agreement in this language. The notable absence of such a development anywhere else in West Iranian (i.e. from clitic pronouns cognate with Table 2 above to object agreement marker) make this a very unlikely scenario – unless one is committed to a cyclic view of the development of agreement for both subject and object pronouns.

## References

- Adibifar, Shirin. 2016. Multi-CAST Persian. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. <https://multicast.aspra.uni-bamberg.de/#persian> (25 June, 2019).
- Baker, Mark. 2011. When agreement is for number and gender but not person. *Natural Language and Linguistic Theory* 29. 875–915. <https://doi.org/10.1007/s11049-011-9147-z>
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge: MIT Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- 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*, 83–100. Wiesbaden: Reichert.
- Fonseca-Greber, Bonnibeth. 2000. The change from pronoun to clitic and the rise of null subjects in spoken Swiss French. Tucson, AZ: University of Arizona dissertation.
- Fuß, Eric. 2005. *The rise of agreement*. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.81>

- Gelderen, Elly van. 2011a. The grammaticalization of agreement. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of grammaticalization*.
- van Gelderen, Elly. 2011b. *The linguistic cycle. Language change and the language faculty*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199756056.001.0001>
- Givón, Talmy. 1976. Topic, pronoun and grammatical agreement. In Charles Li (ed.), *Subject and topic*, 151–188. New York: Academic.
- Hadjidas, Harris & Maria Vollmer. 2015. Multi-CAST Cypriot Greek. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts*. <https://multicast.aspra.uni-bamberg.de/#cypgreek> (25 June, 2019).
- Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A construction grammar approach*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110198614>
- Haig, Geoffrey. 2017. Deconstructing Iranian ergativity. In Jessica Coon, Diane Massam & Lisa Travis (eds.), *The Oxford handbook of ergativity*, 465–500. Oxford: Oxford University Press.
- Haig, Geoffrey. 2018a. Grammaticalization and inflectionalization in Iranian. In Heiko Narrog & Bernd Heine (eds.), *Grammaticalization from a typological perspective*, 57–78. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198795841.003.0004>
- Haig, Geoffrey. 2018b. The grammaticalization of object pronouns: Why differential object indexing is an attractor state. *Linguistics* 56(4). 781–818. <https://doi.org/10.1515/ling-2018-0011>
- Haig, Geoffrey & Širin Adibifar. 2019. Referential Null Subjects (RNS) in colloquial spoken Persian: Does speaker familiarity have an impact? In Alireza Korangy & Behrooz Mahmoodi-Bakhtiari (eds.), *Essays on the typology of Iranian languages*, 102–121. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110604443-006>
- Haig, Geoffrey & Hanna Thiele. 2015. Multi-CAST Northern Kurdish. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST Multilingual Corpus of Annotated Spoken Texts*. <https://multicast.aspra.uni-bamberg.de/#nkurd> (25 June, 2019).
- Jügel, Thomas. 2015. *Die Entwicklung der Ergativkonstruktion im Alt- und Mittliranischen: Eine korpusbasierte Untersuchung zu Kasus, Kongruenz und Satzbau*. Wiesbaden: Harrassowitz. <https://doi.org/10.2307/j.ctvc770qq>
- Kahnemuyipour, Arsalan. 2003. Syntactic categories and Persian stress. *Natural Language and Linguistic Theory* 21. 333–379. <https://doi.org/10.1023/A:1023330609827>
- van Kemenade, Ans. 2000. Jespersen's Cycle revisited: Formal properties of grammaticalization. In Susan Pintzuk, George Tsoulas & Anthony Warner (eds.), *Diachronic syntax: Models and mechanisms*. Oxford: Oxford University Press.
- Korn, Agnes. 2009. Western Iranian pronominal clitics. *Orientalia Suecana* 58. 159–171.
- Lazard, Gilbert. 2006 [1957]. *Grammaire du persan contemporain. Nouvelle édition avec la collaboration de Yann Richard, Rokhsareh Hechmati et Pollet Samvelian*. Téhéran: Institut Français de Recherche en Iran. [Originally published 1957.]
- Lecoq, Pierre. 1979. *Le dialecte de Sivand*. Wiesbaden: Reichert.
- MacKenzie, David. 1961. *Kurdish dialect studies*, vol. 1. Oxford: Oxford University Press.
- MacKenzie, David. 1962. *Kurdish dialect studies*, vol. 2. Oxford: Oxford University Press.
- MacKenzie, David. 1966. *The dialect of Awroman (Hawrāmān-i Luhōn). Grammatical sketch, texts, and vocabulary*. København: Det Kongelige Danske Videnskabernes Selskab.
- Öpengin, Ergin. 2013. Clitic/affix interactions: A corpus-based study of person marking in the Mukri variety of Central Kurdish. Paris III/Bamberg dissertation.
- Öpengin, Ergin. 2016. *The Mukri variety of Central Kurdish: Grammar, texts and lexicon*. Wiesbaden: Reichert.



- Öpengin, Ergin. 2019. Accounting for the combinations of clitic and affix person markers in Central Kurdish. In Gündoğdu, Songül, Ergin Öpengin, Geoffrey Haig & Erik Anonby (eds.), *Current Issues in Kurdish Linguistics*, 243–261. Bamberg: University of Bamberg Press.
- Öpengin, Ergin & Masoud Mohammadirad. Forthcoming. Formal and functional variation in pronominal clitics across Kurdish: documentation and explanations. In Yaron Matras, Ergin Öpengin & Geoffrey Haig (eds.), *Structural and typological variation in the dialects of Kurdish*. London: Palgrave Macmillan.
- Paul, Ludwig. 1998. *Zazaki. Grammatik und Versuch einer Dialektologie*. Wiesbaden: Reichert.
- Rasekh, Mohammad. 2014. Persian clitics: Doubling and agreement. *Journal of Modern Languages* 24(1). 16–33. [http://umijms.um.edu.my/filebank/published\\_article/7262/Rasekh%20-%20JML24%20-%2016-33.pdf](http://umijms.um.edu.my/filebank/published_article/7262/Rasekh%20-%20JML24%20-%2016-33.pdf) (10 October, 2017).
- Samvelian, Pollet. 2007. What Sorani Kurdish absolute prepositions tell us about cliticization. In Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu & Jessica White (eds.), *Proceedings of the Texas Linguistic Society IX: The morphosyntax of understudied languages*, 265–285. Stanford, Calif.: CSLI online publications. [http://web.stanford.edu/group/cslipublications/cslipublications/TLS/TLS9-2005/TLS9\\_Samvelian\\_Pollet.pdf](http://web.stanford.edu/group/cslipublications/cslipublications/TLS/TLS9-2005/TLS9_Samvelian_Pollet.pdf) (accessed 10 October 2017).
- Schiering, René. 2005. Cliticization and the evolution of morphology: A cross-linguistic study on phonology in grammaticalization. Konstanz: University of Konstanz dissertation.
- Schnell, Stefan. 2018. Whence subject-verb agreement? Investigating the role of topicality, accessibility, and frequency in Vera'a texts. *Linguistics* 56(4). 735–780. <https://doi.org/10.1515/ling-2018-0010>
- Siewierska, Anna. 1999. From anaphoric pronoun to grammatical agreement marker: Why objects don't make it. *Folia Linguistica* 33(1–2). 225–251.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511812729>
- Stilo, Don. 2018. 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*, 659–824. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110421682-019>
- Torres Cacoullous, Rena & Catherine Travis. 2019. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57(3). 653–692. <https://doi.org/10.1515/ling-2019-0011>

# The suffix that makes Persian nouns unique

Masoud Jasbi

Stanford University

Although it is widely acknowledged that Tehrani Persian (often broadly labeled as Persian) has no dedicated marker of definiteness, the nominal suffix *-e* has been analyzed as a colloquial definiteness marker. Here I show that *-e* can mark bare nominals to ensure a definite interpretation, but it can also appear on indefinites marked by the indefinite determiner *ye*. I show that indefinites marked by *-e* are scopally inert. To unify the effect of *-e* on definites and indefinites, I propose that *-e* introduces a uniqueness implication on the nominal it modifies. More specifically, *N-e* denotes a singleton set of objects. On a bare nominal, this uniqueness implication ensures a definite interpretation. On an indefinite, it restricts the domain of quantification to a singleton, making the indefinite scopally inert. I present a compositional account of definite and indefinite constructions with *-e* in Tehrani Persian.

**Keywords:** definiteness, specificity, Persian, colloquial

## 1. Introduction

The goal of this chapter is to provide a unified account for the semantics of the nominal suffix *-e* on definite and indefinite constructions in Tehrani colloquial Persian. There is no marker of definiteness similar to the English *the* in Tehrani Persian. Instead, definite descriptions are conveyed using two constructions: simple definites and specific definites. Simple definites are bare nominals that receive a definite interpretation due to implicit contextual cues that support such an interpretation. Specific definites are bare nominals that are modified by the nominal suffix *-e*, and consequently enforce a definite interpretation explicitly.

While the role of *-e* in enforcing a definite interpretation on bare nominals has been discussed before (Ghomeshi 2003), its role on Persian indefinites has remained largely unstudied. Similar to English, Tehrani Persian has an indefinite determiner. I call constructions marked by the indefinite determiner *ye* simple indefinites. The suffix *-e* can also modify simple indefinites, resulting in a construction

which I call specific indefinite. Table 1 provides a summary of the constructions described here, as well as their forms and examples

**Table 1.** Four constructions discussed in this chapter with examples

Construction	Form	Example
Bare Nominal	N	<i>māšīn</i> (“car”)
Specific Definite	N- <i>e</i>	<i>māšīn-e</i> (“the car”)
Simple Indefinite	ye N	<i>ye māšīn</i> (“a car”)
Specific Indefinite	ye N- <i>e</i>	<i>ye māšīn-e</i> (“a certain car”)

In order to understand the semantic contribution of *-e*, I compare the constructions with and without this suffix. § 2 compares the bare nominal and the specific definite, showing that *-e* on a bare nominal enforces a definite interpretation. Section 3 compares the simple indefinite and the specific indefinite. It shows that simple indefinites can take variable scope while specific indefinites always take wide scope with respect to other sentential operators. Section 4 argues that indefinites marked by *-e* are only scopally indefinite and not epistemically specific. Section 5 shows that *-e* does not introduce any common ground requirements and finally Section 6 provides a formal and compositional account of the specific definite and specific indefinite constructions in Persian.

## 2. Bare nominal vs. specific definite

In this section I compare the bare nominal construction (N) and the specific definite construction (N-*e*). The bare nominal construction can be interpreted as generic, indefinite, or definite depending on the utterance context. Example (1) shows the bare nominal *māšīn* (“car”) in three different contexts. In (1a), the bare nominal is used in a context that supports a generic interpretation. In (1b) the bare nominal is interpreted like an indefinite and in (1c) it is interpreted similar to a definite description like *the car*.

- (1) a. [Context: Amir is discussing cars and their problems. He says:]  
*māšīn havā-ro ālude mi-kon-e*  
 car air-OM polluted IPFV-do-3SG  
 “Cars pollute the air.”
- b. [Context: Amir is crossing the street without checking the traffic. Leila stops him and says:]  
*māšīn mi-zan-e be-het*  
 car IPFV-hit-3SG to-2SG  
 “Some car is gonna hit you.”

- c. [Context: Amir and Leila have one car only. One day Amir comes home and says:]  
*māšin xarāb shod-e*  
 car broken become.PST-3SG  
 “The car’s broken.”

What I do next is add the nominal suffix *-e* to each of these sentences and see its effect on the interpretation of the sentences. Example (2) below adds the suffix *-e* to the sentences in Example (1). First, in (2a) I have added *-e* to the generic sentence in (1a). The resulting specific definite construction is not acceptable in a generic context anymore, but it would be acceptable in a new context where *Amir* is referring to a unique car in the utterance context. The addition of *-e* to the bare nominal results in a definite interpretation of the nominal. Next, in (2b) I have added *-e* to the bare nominal in sentence (1b). This new sentence is no longer appropriate for the original context of (1b) and is better suited for a context where a particular car has been introduced such as the one in (2b). It is of course possible to imagine a particular car in the context of (1b) to make the context compatible with the specific definite used. The important intuition is that the interpretation of the specific definite relies on the presence of a unique car in the context. Finally, in (2c) I add *-e* to the sentence in (1c) where the context supported a definite reading. My intuition is that *-e* is completely appropriate for the original context in (1c) and does not alter the original interpretation much; possibly only adding to the salience of the car in the conversation.

- (2) a. [#Context 1a: Amir is discussing cars and their problems. He says:]  
 [Context: Amir shows the video of an old car with a smoky exhaust. He says:]  
*māšin-e havā-ro ālude mi-kon-e*  
 car air-OM polluted IPFV-do-3SG  
 “The/that car pollutes the air.”
- b. [#Context 1b: Amir is going to cross the street without checking the traffic. Leila stops him and says:]  
 [Context: Amir is walking in a parking lot. A car is backing out. Leila stops him and says:]  
*māšin-e mi-zan-e be-het*  
 car IPFV-hit-3SG to-2SG  
 “The/that car is gonna hit you.”
- c. [Context: Amir and Leila have one car only. One day Amir comes home and says:]  
*māšin-e xarāb shod-e*  
 car broken become.PST-3SG  
 “The/that car’s broken.”

The comparison of the examples and contexts in (1) and (2) suggests that the nominal suffix *-e* enforces a unique instantiation of the nominal in the utterance context. In (1) the first two Examples (1a) and (1b) were not interpreted in a context where a unique car was being discussed and adding the suffix *-e* in (2a) and (2b) required new contexts where a unique car was under discussion. The third Example (2c) already had a context with a unique car in the discourse and as a result the addition of the suffix *-e* was compatible with it. These examples suggest that *-e* adds a uniqueness implication to the bare nominal.

### 3. Simple indefinite vs. specific indefinite

In this section I investigate the semantic effect of *-e* on indefinites by comparing the simple indefinite construction (*ye N*) with the specific indefinite construction (*ye N-e*). Example (3) shows the simple indefinite and the specific indefinite constructions in an existential sentence. My judgement is that the example without *-e* in (3a) and the one with *-e* in (3b) receive similar interpretations.

(3) [Context: Leila looks out the window. She says:]

- a. *ye zan dam-e dar-e*  
 ID woman close-EZ door-3SG  
 “A woman is at the door.”
- b. *ye zan-e dam-e dar-e*  
 ID woman-UM close-EZ door-3SG  
 “A woman is at the door.”

The interpretations diverge, however, when we introduce quantificational elements. In (3) I test the scope interaction of the simple and specific indefinite constructions with the universal quantifier *hame*. Tehrani colloquial Persian has two universal quantifiers: *hame* and *har*. *hame* shares some features with the English quantifier *all*. For example they are both used in partitive constructions like ‘all of the students’. *har* is closer to the English quantifiers *every* and *each*. I leave a proper analysis of these universal quantifiers for future work. Here I use *hame* in my examples but the conclusions hold if *hame* is replaced, *mutatis mutandis*, with *har*.

The sentence in (4a) uses the simple indefinite *ye ostād* ‘a professor’ and has two interpretations. First, one in which the universal scopes over the indefinite: for everyone there was a possibly different professor. Second, one in which the indefinite scopes over the universal: everyone said hello to the same professor. In (4b) I have added the suffix *-e* to the indefinite. The only available interpretation in this example is one where the existential scopes over the universal. Therefore, adding *-e* to the indefinite resulted in the indefinite taking wide scope with respect to the universal quantifier.

- (4) a. *emrooz hame be ye ostād salām kard-im*  
 today all to ID professor hello do-1PL  
 “Today we all said hello to a professor.” (1.  $\forall > \exists$ , 2.  $\exists > \forall$ )
- b. *emrooz hame be ye ostād-e salām kard-im*  
 today all to ID professor-UM hello do-1PL  
 “There is a professor that today we all said hello to.” ( $\exists > \forall$ )

What if we have two universal quantifiers? Is it possible to have a simple indefinite scope between the two universal quantifiers? What happens when we add *-e*? In (5) below I construct an example with two universal quantifiers. In (5a) I use a simple indefinite. The example has at least two prominent interpretations: one where the indefinite scopes over the universal quantifiers (the girls corrected the mistakes of the same boy) and one where the indefinite scopes between the universal quantifiers (for every girl there was a different boy whose mistakes were corrected). In (5b) I add the suffix *-e* to the indefinite *ye pesar* “a boy” and the resulting specific indefinite makes only one of the readings available: the one with the indefinite scoping over both universal quantifiers. This example suggests that indefinites with *-e* take the widest scope when interacting with multiple quantifiers.

- (5) a. *hame-ye doxtar-ā hame-ye eštabā-hā-ye ye pesar ro*  
 all-EZ girl-PL all-EZ mistake-PL-EZ ID boy OM  
*tasih kard-an*  
 correct did-3PL  
 “All the girls corrected all the mistakes of a boy.” (1.  $\exists > \forall > \forall$  2.  $\forall > \exists > \forall$ )
- b. *hame-ye doxtar-ā hame-ye eštabā-hā-ye ye pesar-e ro*  
 all-EZ girl-PL all-EZ mistake-PL-EZ ID boy-UM OM  
*tasih kard-an*  
 correct did-3PL  
 “There is a boy that every girl corrected all his mistakes.” ( $\exists > \forall > \forall$ )

In (6), *-e* shows a similar wide-scope effect with respect to the temporal quantifier *hamiše* “always”. In (6a) I use a simple indefinite which allows two interpretations: one with the existential claim scoping over the temporal quantifier (it’s always the same boy) and one where the existential scopes below the temporal quantifier (it’s a different boy every time). Adding *-e* in (6b) only allows the second reading with the indefinite taking wide scope.

- (6) a. *Sārā hamiše bā ye pesar davā-š mi-š-e*  
 sara always with ID boy quarrel-3SG IPFV-become-3SG  
 “Sara always gets into a fight with some boy.” (1.  $\exists > \text{always}$  2.  $\text{always} > \exists$ )
- b. *Sārā hamiše bā ye pesar-e davā-š mi-š-e*  
 sara always with ID boy quarrel-3SG IPFV-become-3SG  
 “Sara always gets into a fight with some boy.” ( $\exists > \text{always}$ )

Next in (7), I test the simple and specific indefinite constructions in the *de-rel de-dicto* contexts. In (7a) I use the simple indefinite and the sentence allows two interpretations: one with the indefinite scoping over the modal ‘want’ (there is a specific girl) and one with the indefinite scoping under the modal (Amir wants to marry just any girl). Yet again adding the suffix *-e* only allows the wide scope existential reading as (7b) shows.

- (7) a. *Amir mi-xā-d bā ye doxtar ezdevāj kon-e*  
 amir IPFV-want-3SG with ID girl marriage do-3SG  
 “Amir wants to marry a girl.” (1.  $\exists > \text{want}$  2.  $\text{want} > \exists$ )
- b. *Amir mi-xā-d bā ye doxtar-e ezdevāj kon-e*  
 amir IPFV-want-3SG with ID girl marriage do-3SG  
 “There is a girl Amir wants to marry.” ( $\exists > \text{want}$ )

In (8) I look at the scope relation of the simple and specific indefinite constructions with the belief verb *fekr kardan* ‘to think’. The simple indefinite in (8a) has at least two prominent interpretations. On the first interpretation, there is a unique girl that everyone thinks Ali has married. On the second interpretation, everyone thinks that Ali has married a girl but they may think of different girls (e.g. Ali thinks Amir has married Targol but Hasan thinks Amir has married Leila.) In (8b) I have used the specific indefinite *ye doxtar-e* ‘a girl-e’ and the only available interpretation is the one in which everyone is thinking of a specific girl that Amir has married. This example shows that the indefinite marked by *-e* takes the widest scope even in the presence of a universal quantifier and a belief verb.

- (8) a. *hame fekr mi-kon-an Ali bā ye doxtar ezdevāj kard-e*  
 all thought IPFV-do-3PL ali with ID girl marry do-PERF.3SG  
 “Everyone thinks Ali has married a girl.” (1.  $\exists > \forall > B$  2.  $\forall > B > \exists$ )
- b. *hame fekr mi-kon-an Ali bā ye doxtar-e ezdevāj kard-e*  
 all thought IPFV-do-3PL ali with ID girl-UM marry do-PERF.3SG  
 “There is a girl everyone thinks Ali has married.” ( $\exists > \forall > B$ )

Finally I test the behavior of *-e* in the antecedent of conditionals marked by *age* ‘if’. The simple indefinite example in (9a) has two prominent interpretations: first that Amir will be happy if he marries a specific girl (e.g. Leila) and second, that Amir will be happy if he marries any girl. As expected, in (9b) where the suffix *-e* is present on the noun, the only available interpretation is the specific one: there is a specific girl that if Amir marries, he will be happy.

- (9) a. *age Amir bā ye doxtar ezdevāj kon-e, xeili xošhāl mi-še*  
 if amir with ID girl marriage do-3SG, very happy IPFV-become.3SG  
 “If Amir marries a girl, he will be very happy.” (1.  $\exists > \text{if}$  2.  $\text{if} > \exists$ )

- b. *age Amir bā ye doxtar-e ezdevāj kon-e, xeili xošhāl mi-še*  
 if amir with ID girl-UM marriage do-3SG, very happy IPFV-become.3SG  
 “There is a girl that if Amir marries, he will be happy.” (∃ > if)

To summarize, in this section I compared the simple indefinite (*ye N*) and the specific indefinite (*ye N-e*) constructions and showed that the presence of the suffix *-e* on an indefinite systematically picks the widest scope for that indefinite. Crucially, from the brief but relatively wide array of quantificational and scope taking items used in this section, it appears that the wide-scope tendency of the specific indefinite is very strong and independent of the nature and number of the other operators involved.

#### 4. Specificity

Farkas (1994) discusses three types of specificity: epistemic, scopal, and partitive. The investigation of the specific indefinite in the previous section suggests that indefinites with *-e* in Tehrani Persian are scopally specific. Since scopally specific indefinites may also be epistemically specific, here I investigate whether the Persian indefinites with *-e* are also epistemically specific; meaning the speaker has a specific referent in mind when uttering them. It is important to emphasize that the issue here is not whether speakers can have a specific referent in mind when using indefinites with *-e*. They certainly can and many examples in the previous sections can show this. The issue is whether speakers must necessarily have a specific referent in mind when they use an indefinites marked by *-e*. In other words, are all indefinites with *-e* epistemically specific?

The examples below in (10) show that the answer is no. (10a) is a naturally occurring example from twitter. It is not at all necessary for the speaker or the addressee to know who the girl in this example is. In fact, the context makes it likely that the speaker did not know the girl his friend was chatting with. Similarly in (10b), the specific indefinite can be uttered to convey the news that some man has committed suicide but there is no need for the speaker or the addressee to know exactly who this man is. This is similar to the usage of *some man* or *a certain man* in English. Examples like the ones in (10) suggest that the specific indefinite construction (*ye N-e*) is not epistemically specific.

- (10) a. *dust-am eštabāhi eskirin-šāt-e čat-eš-o bā*  
 friend-1SG mistakenly screen-shot-EZ chat-3SG-OM with  
*ye doxtar-e fereštād*  
 ID girl-UM sent.3SG  
 “My friend mistakenly sent me a screen shot of his chat with a girl.”



- b. *mi-g-an tu Shiraz ye mard-e xod-koši kard-e*  
 IPFV-say-3PL in Shiraz ID man-UM self-killing did-PERF.3SG  
 “They say a man has committed suicide in Shiraz.”

## 5. Common ground effects

In this section I compare the specific definite (N-*e*) and the specific indefinite (*ye* N-*e*) constructions with respect to their effect on the common ground; i.e. the mutual public knowledge between the speaker and the addressee in the discourse. Consider two families: the Tehrani family and the Yazdi family. Suppose that we know that the Tehrani family has only one son. We don’t know anything about the Yazdi family. Now looking at the examples in (11), (11a) can be said felicitously to convey that the son in the Tehrani family is married. It provides further information about the unique son in the Tehrani family we knew about. However, the sentence in (11b) about the Yazdi family is not as felicitous. It acts as if we knew about a son in the Yazdi family when we did not. There is a sense of imposing further information or asking the listener to accommodate information about the Yazdi family that was not in common ground before.

- (11) a. In the Tehrani family, ...  
*pesar-e ezdevāj kard-e*  
 son-UM marriage do-PERF.3SG  
 “The son has married.”
- b. In the Yazdi family, ...  
 #*pesar-e ezdevāj kard-e*  
 son-UM marriage do-PERF.3SG  
 “The son has married.”

Compare the previous example with the ones in (12) where I use a specific indefinite instead of the specific definite. The addition of the indefinite determiner *ye* flips the felicity judgments. It is not felicitous to use the specific indefinite in (12a) to talk about the son in the Tehrani family. The reason is that we already know about the son and a definite serves the reference to the son better than an indefinite. However, it is completely felicitous to use the specific indefinite in (12b) to inform the listener about a son in the Yazdi family. The specific indefinite is suitable for introducing new information. The examples here show that even though the suffix *-e* appears on definites and indefinites, it has no role in determining the familiarity of the nominal. Familiarity is controlled by the presence or absence of the indefinite determiner *ye*.

- (12) a. In the Tehrani family, ...  
 #*ye pesar-e ezdevāj kard-e*  
 ID SON-UM marriage do-PERF.3SG  
 “A son has married.”
- b. In the Yazdi family, ...  
*ye pesar-e ezdevāj kard-e*  
 ID SON-UM marriage do-PERF.3SG  
 “A son has married.”

## 6. Analysis

Summarizing the findings of the previous sections, in § 2, I showed that the presence of *-e* on a bare nominal enforces a definite interpretation. In § 3, I showed that the presence of *-e* on an indefinite forces it to take the widest scope with respect to the other sentential operators. In § 4, I argued that indefinites with *-e* are not epistemically specific and in § 5 I showed that *-e* does not place any requirements on the common ground. The goal in this section is to unify these observations and propose a single lexical entry for *-e* that captures these effects.

I propose that the nominal suffix *-e* encodes the uniqueness of the nominal in the utterance context. To use an example, *māšin-e* (“car”-*e*) conveys that there is a unique car in the utterance context. This proposal captures the empirical observations in § 2 and § 3. With definites, the uniqueness implication introduced by *-e* is an essential part of the definite description (Russell 1905; Abbott 2006). On an indefinite, the uniqueness implication results in a singleton indefinite (Schwarzschild 2002). The singleton indefinite is scopally inert; it does not participate in scope interactions and gives the impression of wide scope.

There are two more observations that I would like to capture in my analysis here. First, the uniqueness implication of *-e* is not affected by entailment canceling operators such as the antecedent of conditionals. Examples such as (9) suggest that the uniqueness implication of *-e* escapes the influence of entailment canceling operators and it is enforced globally. The interpretation of (9) is not ‘if there is a unique girl that Amir marries, he will be happy’. In other words, the consequent does not depend on the uniqueness implication of the antecedent. The existence and uniqueness of ‘girl’ is interpreted outside the scope of the conditional: ‘there is a unique girl and if Amir marries the girl, he will be happy’. This observation suggests that even though the uniqueness implication is introduced in the antecedent by *-e*, it should be passed up the derivation tree unaffected by entailment canceling operators until it is interpreted globally. Second, the contribution of *-e* is not presuppositional. In § 5, I showed that the usage of *-e* does not require a common

ground that presupposes the uniqueness of the nominal description. The specific indefinite with *-e* can be felicitously used to introduce new information. The first and the second observations suggest that even though the uniqueness implication of *-e* is projective, it is not presuppositional. To capture these two observations in my analysis, I treat the uniqueness implication of *-e* as a conventional implicature using Potts (2005)'s two-dimensional system. This way we can guarantee that the uniqueness implication is always enforced globally.

Figure 2 shows sample derivation trees for simple definite and indefinite constructions in Tehrani colloquial Persian. A bare nominal that picks out a unique entity in the utterance context can be covertly type-shifted via Partee (1986)'s *iota* operator. On the other hand, a simple indefinite like *ye māšīn* is similar to 'a car' in English. The indefinite determiner *ye* introduces an existential quantifier.

Figure 3 shows the derivation of an example specific indefinite. The black dot separates the at-issue or ordinary content of the sentence (left) from the projective content (right). A specific indefinite is derived similar to a simple indefinite, except that a uniqueness implication is introduced by the nominal suffix *-e* and passed up in the projective dimension of the tree. This uniqueness implication will not be affected by other sentential operators and will be interpreted globally, ensuring that the indefinite will be scopally inert.

Figure 3 shows the derivation of a sample specific definite construction. The derivation of a specific definite is similar to that of a simple definite shown in Figure 2. The main difference is that similar to a specific indefinite, a uniqueness implication is introduced by the nominal suffix *-e* which is passed up the tree as projective content. Since the nominal is marked explicitly as unique and the indefinite determiner is absent, the nominal is again type-shifted by *iota*. We can say that in specific definites in Persian, *-e* does explicitly what the context of the utterance often does implicitly with bare nominals: ensure that the nominal denotes a unique entity in the utterance context.

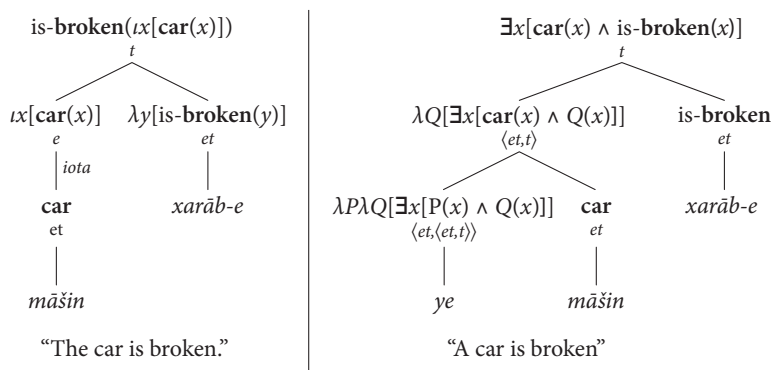


Figure 1. Derivations for sample definite and simple indefinite constructions in Persian

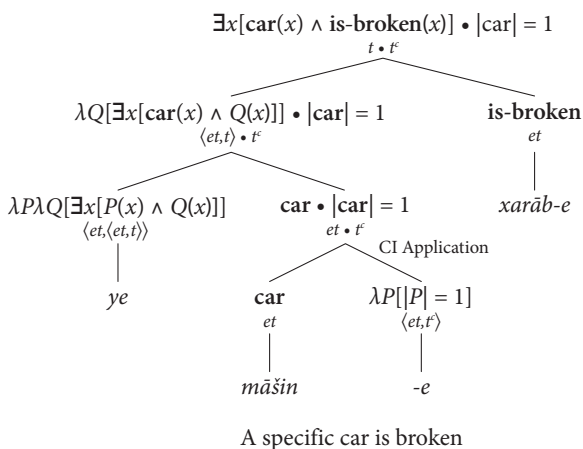


Figure 2. Derivation for a sample specific indefinite construction in Persian

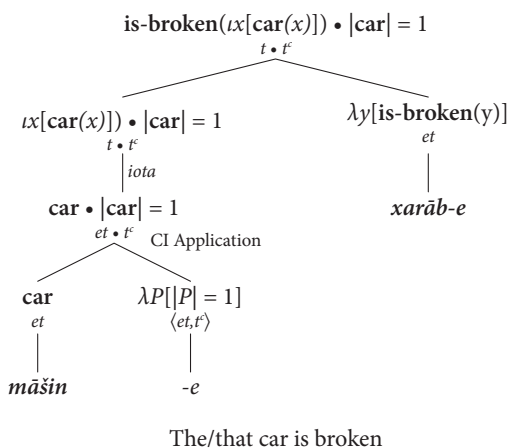


Figure 3. Derivation for a sample specific definite construction in Persian

## 7. Conclusion

I investigated the semantics of four nominal constructions in Tehrani colloquial Persian: bare nominals (N), simple indefinites (*ye* N), specific indefinites (*ye* N-*e*), and specific definites (N-*e*). I first confirmed previous reports that bare nominals in Persian can be interpreted as either generic, indefinite, or definite, depending on the utterance context (Toosarvandani and Nasser 2015). When the utterance context supports a definite reading, simple definite descriptions such as ‘the car’ surface as bare nominals. I showed that specific definites are similarly suitable for

contexts that support definite interpretations. However, by adding the suffix *-e*, the specific definite enforces a definite reading regardless of the context. In other words, in Tehrani Persian, *-e* does explicitly what utterance context often does implicitly.

Second, a simple indefinite (e.g. *ye māšīn*) shows similar scope taking properties to a simple indefinite in English (e.g. *a car*). I have shown that adding the suffix *-e* to a simple indefinite (e.g. *ye māšīn-e*) results in a scopally specific indefinite: the indefinite takes the widest possible scope with respect to sentential operators. Finally, to provide a unified account for the semantic contribution of *-e* in definite and indefinite constructions, I have proposed that *-e* carries a uniqueness implication and requires the nominal to denote a singleton set. I have provided examples that suggested the uniqueness implication of *-e* is projective but not presuppositional, and presented a formal account that captured the empirical observations discussed in this chapter.

## References

- Abbott, Barbara. 2006. Definite and indefinite. *Encyclopedia of Language and Linguistics* 3. 392–399. <https://doi.org/10.1016/B0-08-044854-2/01089-0>
- Farkas, Donka F. 1994. Specificity and scope. In Lea Nash & George Tsoulas (eds.), *Proceedings of Langues et Grammaire*, vol. 1, 119–137. Paris: University of Paris 8.
- Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia Linguistica* 57(2). 47–74. <https://doi.org/10.1111/1467-9582.00099>
- Partee, Barbara. 1986. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jongh, & Martin Stokhof (eds.), *Studies in discourse representation theory and the theory of generalized quantifiers*, 115–143. Dordrecht: Foris.
- Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.
- Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493. <https://doi.org/10.1093/mind/XIV.4.479>
- Schwarzschild, Roger. 2002. Singleton indefinites. *Journal of Semantics* 19(3). 289–314. <https://doi.org/10.1093/jos/19.3.289>
- Toosarvandani, Maziar and Hayedeh Nasser. 2015. Quantification in Persian. In Edward Keenan & Denis Paperno (eds.), *Handbook of quantifiers in natural language*, 2nd edn. Dordrecht: Springer.

# The meaning of the Persian object marker *rā*

## What it is not, and what it (probably) is

Masoud Jasbi

Stanford University

The Persian object marker *rā* is called many things, among them: marker of specificity (Karimi 1990), definiteness (Mahootian 1997), secondary topics (Dabir-Moghaddam 1992), and presuppositions (Ghomeshi 1996). These accounts capture the core of what *rā* is, yet also include a lot of what *rā* is NOT. I report novel examples that show *rā* is NOT an (exclusive) marker of specific or definite referents. It is also NOT an (exclusive) marker of (secondary) topics. Instead, *rā*'s core contribution is something shared by all these accounts: old or presupposed information. I show that the information presupposed by *rā* is an existence implication. A marked object like *sandali-ro* ("chair" -*rā*) implies that there is one or more chairs in the conversational context. This account captures several novel observations on the distribution of *rā* such as its optional presence on proper names in some contexts. I provide a formal and compositional analysis of simple Persian sentences with definite and indefinite objects.

**Keywords:** *rā*, specificity, definiteness, presupposition, topic, information structure

### 1. Introduction

The object marker *rā* is a familiar topic in Iranian linguistics. Previous proposals have described its function as SPECIFICITY marking (Karimi 1990, 1999, 2003), TOPIC marking (Dabir-Moghaddam 1992; Dalrymple & Nikolaeva 2011), DEFINITENESS marking (Mahootian 1997), and PRESUPPOSITION marking (Ghomeshi 1996). The goal of this chapter is to first show that even though these accounts differ in some respects, they all share the core theoretical notion of PRESUPPOSED INFORMATION. Second, the chapter aims to develop this shared insight and provide a formal and compositional account of object marking in Tehrani Persian. Here is what I do: first, I provide novel examples that show where some of the previous

accounts have incorrect predictions (§ 2); second, I argue that *rā* marks existential presuppositions (§ 3); third, I provide a formal and compositional account of Persian simple definite and indefinite constructions with *rā* (§ 4).

## 2. What *rā* is not

In this section, I discuss the theoretical notions that do not accurately capture the semantic contribution of the object marker. These notions include secondary topics, specificity, and definiteness. For each notion, I first discuss what I mean by it and then provide examples that show the notion does not quite fit the distribution of *rā*.

### 2.1 Topics

Information structural accounts hypothesize a direct mapping between pragmatics and syntax (Dabir-Moghaddam 1992; Dalrymple & Nikolaeva 2011). In such accounts, the sentence is divided into two parts: TOPIC and FOCUS. Topic contains old information (presupposition) and focus contains new information (assertion). There can be two topics: primary and secondary. Nikolaeva (2001) defines ‘primary topic’ as the entity that the sentence is about. ‘Secondary topic’ is defined as an entity such that the sentence is construed to be ABOUT its relationship with the primary topic. Not every sentence has a secondary topic but every sentence has a primary topic and a focus. All elements in the sentence are assigned to be topic or focus and no element can be both. In (1), the parts of the utterances that are primary topic, secondary topic, and focus are shown within brackets subscripted as T1, T2, and f respectively.

- (1) a. Whatever became of John? (Dalrymple & Nikolaeva 2011)  
 b. [He]<sub>T1</sub> [married Rosa]<sub>f</sub>.  
 c. but [he]<sub>T1</sub> [didn’t really love]<sub>f</sub> [her]<sub>T2</sub>

Dabir-Moghaddam (1992) proposed that *rā* marks secondary topics in the sense defined above. The proposal is accurate in that it highlights the notion of familiarity and hypothesizes that *rā* carries presuppositional content. However, the information structural framework seems too rigid to capture the distribution of *rā* properly. It fails to capture examples where *rā* appears on nominals that are primary topic or focus. For example in (2), the object marker appears on primary topics.

- (2) a. [čī]<sub>f</sub> [John]<sub>T1</sub> o [košt]<sub>f</sub>?  
 what John OM kill.PST.3SG  
 “What killed John?”

- b.  $[ki]_f$   $[ye\ m\dot{a}š\dot{in}]_{TI}$  *o*  $[dozdid]_f?$   
 who ID car OM steal.PST.3SG  
 “Who stole a car (one of the cars)?”

Examples in (3) show that the object marker can also appear on question particles that are traditionally analyzed as focus and carry new information.

- (3) a.  $[Amir]_{TI}$   $[ki]_f$  *ro*  $[did]_f?$   
 Amir who OM see.PST.3SG  
 “Who did Amir see?”  
 b.  $[Amir]_{TI}$   $[kodum\ keik]_f$  *o*  $[xord]_f?$   
 Amir which cake OM eat.PST.3SG  
 “Which cake did Amir eat?”

More importantly, research in formal semantics and pragmatics suggests that the same lexical item can contribute both presupposed and asserted content. This is the case with some presupposition triggers such as ‘stop’, ‘continue’, and ‘only’. It may be possible to tweak the information structural accounts to address these problems but I believe there are already available tools in formal semantics that can help us better capture the meaning of *rā*. This is the path I pursue in § 4.

## 2.2 Specificity

Farkas (1994) defines three types of specificity: epistemic, scopal, and partitive. In this sub-section I discuss each and show that the first two do not capture the semantic contribution of *rā* while the last one does. However, it is hard to call partitive specificity ‘specific’ in a proper sense. So I advocate against using the term ‘specificity’ to describe the meaning of *rā*.

### *Epistemic specificity*

An NP is “epistemically specific” if it denotes a specific (unique) entity that the speaker has in mind (Fodor & Sag 1982). In (4) below, the continuations (4a) and (4b) clarify the epistemically specific and nonspecific readings respectively. In the epistemically specific reading, the speaker knows the specific referent of ‘a movie’. In the epistemically nonspecific reading, the speaker does not know the specific referent of ‘a movie’. In either case, ‘a movie’ refers to a unique entity and its value does not vary with the universal quantifier *hame* ‘all’.

- (4) *hame ye film tamāšā kard-an*  
 all ID film watch do.PST-PL  
 “Everyone watched a movie.”



- a. *esm-eš darbāre-ye Eli bud*  
 name-3SG about-EZ Eli be.PST.SG  
 “its name was About Eli.” (Epistemically Specific)
- b. *ne-mi-dun-am či bud*  
 NEG-IPFV-know-1SG what be.PST.SG  
 “don’t know what it was.” (Epistemically nonspecific)

Karimi (1990) defines specificity as “denoting a specific individual”. Under her proposal, specific NPs divide into specific definites and specific indefinites. Specific definites denote individuals that are known to the speaker and the hearer. Specific indefinites denote individuals that are only known to the speaker. This amounts to epistemic specificity explained above.

Here I argue that *rā* does not mark epistemic specificity by showing that: (i) an entity denoted by a *rā*-marked object can be unknown to the speaker and (ii) an entity known to the speaker can appear without *rā*. In other words, *rā*-marked objects can be epistemically nonspecific and epistemically specific indefinites can appear without *rā*. I start with the first leg of the argument. In (5), the context is designed in a way that the referent of the *rā*-marked object is unknown to the speaker (i.e. epistemically nonspecific), yet object marking is obligatory.

- (5) [Context: Ali’s three-year-old child takes his phone and accidentally deletes a picture. He sees the number of pics drop to 99 from 100 but he doesn’t know which picture is deleted. He says:]

*ne-mi-dun-am kodum aks-o in bače pāk kard-e*  
 NEG-IPFV-know-1SG which pic-OM this kid clean do-PST.3SG

“I don’t know which picture this kid has deleted.”

In (6), object marking is optional and whether the object marker is present or not, the prominent reading is not one in which the speaker is requesting a specific plate they have in mind. These examples show that the presence or absence of the object marker is not related to the speaker having a specific referent in mind or not.

- (6) [Context: Ali is at the dinner table. There are some plates on the other end of the table. He asks his brother to give him a plate.]

*ye bošqāb(-o) mi-d-i?*  
 ID plate-OM IPFV-give-2SG

“Can you give me a plate?”

Considering the second leg of the argument, many individuals known to the speaker (i.e. epistemically specific) appear without *rā*. In (7) below, it is clear from the sentence itself that the speaker knows the specific referent of the indefinite NP *ye xune* “a house”. The speaker has visited and seen the house. However, the

epistemically specific NP is appearing comfortably without the object marker. It is quite easy to construct examples like this and they show us that *rā* is not a marker of epistemic specificity.

- (7) *diruz ye xune did-im tu Ferešteh*  
 yesterday ID house see.PST-3PL in Fereshteh  
 “We saw a house in Fereshteh yesterday.”

### Scopal Specificity

I define an indefinite NP as “scopally specific” if it takes wide scope with respect to all other sentential operators (e.g. quantifiers).<sup>1</sup> (8a) and (8b) distinguish the scopally specific and nonspecific readings of the indefinite ‘a movie’ respectively. The scopally specific reading is also known as wide scope existential reading and the nonspecific reading as narrow scope existential reading.

- (8) *hame ye film tamāšā kard-an*  
 all ID film watch do.PST-PL  
 “Everyone watched a movie.”
- a. *ye film-e xās*  
 ID film-EZ specific  
 “a specific movie.” (Scopally specific,  $\exists > \forall$ )
- b. *film-hā-ye moxtalef*  
 film-PL-EZ different  
 “different films.” (Scopally nonspecific,  $\forall > \exists$ )

The proposal in Karimi (1990) also assumes that *rā* marks scopal specificity. This is because epistemic specificity subsumes scopal specificity: if an NP is epistemically specific, then it is scopally specific too. However, not every scopally specific NP is epistemically specific. So if *rā* does not mark epistemic specificity, does it mark scopal specificity? Here I show that (i) *rā* appears on nominals that are not scopally specific; and that (ii) scopally specific indefinites can appear without *rā*.

In (9), the object marker appears on *doxtar* “girl” even though the prominent reading is a scopally nonspecific one in which every boy chose a different girl.

- (9) [Context: Dance class; equal number of girls and boys. Boys have to choose partners.]  
*har pesar-i ye doxtar-o entexāb kard*  
 each boy-IC ID girl-OM choose do.PST.3PL  
 “Every boy chose a (different) girl.” ( $\forall > \exists$ )

1. Farkas (1994)’s definition is slightly different in that she allows the term to refer to indefinites that take wide scope with respect to some (not necessarily all) other operators.

In (10), the indefinite *ye kār* ‘a job’ scopes below the modal and receives a *de dicto* reading, yet it appears with the object marker. Notice that the indefinite in (10) is both scopally and epistemically nonspecific.

- (10) [Context: Maryam has three job offers. She has to pick one by tomorrow.]  
*mi-xād ye kār-o tā fardā qabul kon-e vali hanuz*  
 IPFV-want.3SG ID job-OM until tomorrow accept do.PST-3PL but yet  
*ne-mi-dun-e kodum-o*  
 NEG-IPFV-know-3SG which-OM  
 ‘She wants to accept a job by tomorrow but she still doesn’t know which’  
 (want > ∃)

On the other hand in (11), the indefinite *ye qazā* ‘a food’ scopes out of two universal quantifiers without having the object marker *rā*.

- (11) [Context: A restaurant where everyone always orders burgers. The waiter says:]  
*injā hame hamiše ye qazā sefāreš mi-d-an*  
 here all always ID food order IPFV-do.PST-3PL  
 ‘Everyone always orders the same food here.’ (∃ > ∀ > ∀)

More generally, it is hard to find a correlation between scope and object marking in Persian and in a lot of examples like (12) both wide scope and narrow scope readings are available. I conclude that *rā* is not a marker of scopal specificity either.

- (12) *hame-ye pesar-ā ye doxtar-o dust dār-an*  
 all-EZ boy-PL ID girl-OM friend have.NPST-3PL  
 ‘All the boys love some girl.’ (∀ > ∃)  
 ‘There is a girl that all the boys love.’ (∃ > ∀)

### Partitive specificity

An NP is ‘partitively specific’ if it is interpreted as part of a set introduced in previous discourse (Enç 1991). In (13), since the context introduces the salient set of movies on Netflix, the indefinite ‘a movie’ may have a partitive reading like ‘one of the movies on Netflix’. The continuation in (13a) picks this reading. In the nonspecific reading in (13b), the denotation of ‘movie’ is not restricted to the set of movies on Netflix.

- (13) [Context: After joining Netflix ...]  
*hame ye film tamāšā kard-an*  
 everyone ID film watch do.PST-PL  
 ‘Everyone watched a movie.’  
 a. *ye film tu Netflix*  
 ID film in Netflix  
 ‘a movie on Netflix’ (Partitively specific)

- b. *ye film tu sinemā*  
 ID film in theater  
 “a movie in the movie theater” (Partitively nonspecific)

Karimi (2003) proposes that *rā* marks partitive specificity. This is more or less the semantic characterization of *rā* that I propose below. However, I have reservations about calling *rā* a specificity marker. The main reason is that the term ‘specificity’ is commonly associated with epistemic or scopal specificity in the wider linguistics literature. And as I argued above, these two types of specificity misrepresent what *rā* does semantically.

More importantly, there is a fundamental difference between epistemic/scopal specificity and partitive specificity. Epistemic and scopal specificity rely on the notion of ‘fixed reference’. When the speaker knows the referent of a nominal, the referent is fixed and cannot vary with respect to other operators such as modals or quantifiers. This is why ‘a movie’ takes wide scope in (4a) and only picks ‘about Eli’. Scopal specificity allows variation of referent with respect to the epistemic state of the speaker but does not do so for operators such as quantifiers. Therefore, ‘a movie’ takes wide scope with respect to *hame* in (8a) but for all the speaker knows, the referent may be ‘about Eli’, ‘a separation’, or ‘the salesman’, etc. Fixed reference or better, as Farkas and Brasoveanu (2013) put it, “stability and variability in assignment function” is the essence of specificity.

Partitive specificity does not fit the essence of specificity. It relies on familiarity: old vs. new information. The referent of a partitively specific NP is not necessarily fixed with respect to any operator. In (13a), everyone may have watched the same movie or different ones; if they watched the same movie, the speaker may know the movie or not. Similar to information structural accounts, the core distinction in partitive specificity is familiarity: given vs. new information. Partitive specificity delimits the set of objects that an indefinite like ‘a movie’ can refer to and makes it common ground between discourse participants. This fundamental difference between epistemic/scopal specificity and partitive specificity makes it inaccurate to call *rā* a specificity marker.

While it is possible to call *rā* a marker of partitive specificity, I believe the better option is to avoid specificity altogether and use the notion of presuppositionality to label the meaning of *rā* (Ghomeshi 1996). The notion of ‘presupposition’ has the advantage that it brings together the insights in the information structural account of Dabir-Moghaddam (1992) and the definiteness account of Mahootian (1997) as well. In the next section I define what I mean by ‘presupposition’ more accurately when I elaborate on the notion of definiteness.

### Definiteness

I borrow the key notions of definiteness from the classical accounts of Russell (1905) and Strawson (1950). See Abbott (2006) for a general discussion of these notions. I use the term implication as a general and neutral term to refer to linguistic meaning. An implication may be an entailment, a presupposition, an implicature, or any other type of meaning. I say a nominal implies existence if it denotes a nonempty set ( $|\llbracket[\text{NP}]\rrbracket| \geq 1$ ) in the conversational context. A nominal implies uniqueness if it denotes a singleton set ( $|\llbracket[\text{NP}]\rrbracket| = 1$ ). For example, the nominal *golābi* ‘pear’ implies existence in both (14a) and (14b); it implies that there is at least one pear in the conversational context. However, only in (14a) there is also a uniqueness implication: that there is only one (relevant) pear in the conversational context.

- (14) a. *man golābi ro xord-am*  
 I pear OM eat-1SG  
 ‘I ate the pear.’ (↗ Existence + Uniqueness)
- b. *man ye golābi xord-am*  
 I ID pear eat-1SG  
 ‘I ate a pear.’ (↗ Existence)

Following Stalnaker (1978), I define ‘common ground’ as the mutually recognized shared information between the speaker(s) and the addressee(s). I call an implication ‘presuppositional’ if it is the result of constraints on the common ground. I test presuppositionality by constructing conversational contexts as minimal pairs in which the relevant implication is or is not common ground between discourse participants. If an implication is presuppositional, it is only acceptable when the conversational common ground already includes it. For example, the existence and uniqueness of *golābi* ‘pear’ is not common ground in (15). In such a context, the definite construction in (15a) is unacceptable while the indefinite construction in (15b) is perfectly fine.

- (15) [Context: Mona has bought a pear. Eli is in her room and does not know this. Mona eats the pear, goes to Eli’s room and says:]
- a. # *man golābi ro xord-am*  
 I pear OM eat-1SG  
 ‘I ate the pear.’ (↗ Existence + Uniqueness)
- b. *man ye golābi xord-am*  
 I ID pear eat-1SG  
 ‘I ate a pear.’ (↗ Existence)

In (16) below, the context is minimally changed to make the uniqueness implication of *golābi* ‘pear’ common ground between Mona and Eli. Now the judgments flip: the definite construction in (16b) is perfectly fine while the indefinite construction

in (16a) is odd. The results in (15) and (16) are consistent with the hypothesis that definite constructions presuppose uniqueness of their denotations. Definites are commonly considered to presuppose existence and uniqueness while indefinites carry an existence entailment, and can give rise to anti-uniqueness implications (Coppock & Beaver 2012; Heim 1991).

- (16) [Context: Mona and Eli bought a pear together. Later, Mona eats the pear, goes to Eli's room and says:]
- a. *man golābi rā xord-am*  
 I pear OM eat-1SG  
 "I ate the pear." (↗ Existence + Uniqueness)
- b. # *man ye golābi xord-am*  
 I ID pear eat-1SG  
 "I ate a pear." (↗ Existence)

Establishing an implication as a presupposition takes more than checking its common ground status. We need to also test its projection properties using the family-of-sentences diagnostic (Chierchia & McConnell-Ginet 1990). For a comprehensive discussion of projection, see Tonhauser et al. (2013). Since a full presentation of projection properties for the definite and indefinite constructions in Persian is beyond the scope of this chapter, I refer the reader to Jasbi (2015, 2016) for a more comprehensive treatment.

Mahootian (1997) proposed that *rā* is a marker of definiteness. The main reason for this proposal was contrasts like (17a) and (17b), in which *rā* seems to carry the meaning of the definite article in English. However, it was evident to Mahootian (1997) that a definiteness account cannot capture what *rā* does due to examples like (17c): *rā* commonly appears with the indefinite determiner *ye*. To resolve this issue, Mahootian (1997: 201) defined definiteness as a scale and suggested that *rā* marks object NPs toward the more definite end of the scale. She maintained that NPs like the one in (17c) are "somewhat definite" since "they refer to some delimited class of objects" (Mahootian 1997: 201). Notice that a 'delimited set of objects' is exactly what the notion of partitive specificity captures as well. Therefore, despite terminological differences, the definiteness and partitive specificity accounts of *rā* have a lot in common.

- (17) a. *Ali ketāb xarid*  
 Ali book buy.PST.3SG  
 "Ali bought one or more books."
- b. *Ali ketāb-o xarid*  
 Ali book-OM buy.PST.3SG  
 "Ali bought the book."

- c. *Ali ye ketāb-o xarid*  
 Ali ID book-OM buy.PST.3SG  
 “Ali bought one of the books.” (Partitive)  
 “Ali bought a certain book.” (Epistemic)

In the next section, I argue that *rā* carries an existence presupposition. Given that definite descriptions are often considered to carry existence and uniqueness presuppositions, it is not surprising that *rā* appears with definites. However, since indefinites can also carry existence presuppositions, *rā* can appear on them too. In some ways, Mahootian (1997) was quite accurate to say that *rā*-marked nominals are somewhat definite. They carry half of the presuppositional content of a definite (the existence presupposition). In the next section I make these intuitions more precise.

### 3. What *rā* (probably) is

I propose that *rā*'s semantic contribution has at least two components, first, an existence claim on the NP that it modifies. For example, if it modifies *sandali* “chair”, it implies that the set of objects denoted by *sandali* “chair” is nonempty ( $|\llbracket \text{NP} \rrbracket| \geq 1$ ). This may seem like a trivial implication. Almost all nominals denote nonempty sets of objects so why mark them? This is why the second component is crucial and has attracted most of the attention in the literature. *Rā* also signals that this existential implication is part of the common ground in the conversation. In other words, it is presuppositional. Putting these two pieces together, an NP such as *sandali-ro* (“chair”-*rā*) implies that there is one or more mutually known chairs *in the current conversational context*.

In what follows, I present a few examples that I have found most convincing for the account proposed above.<sup>2</sup> Consider the first component: the existence claim. (18a) and (18b) bellow are minimal pairs; the first does not have the object marker but the second does. Both examples start with a clause that negates the existence of any task or work in the context of the example. In other words, *kār* “work” denotes an empty set in the context of (18). If the object marker requires the nominal to be nonempty, then it should be unacceptable on *kār* “work”. This is exactly what we find comparing (18a) and (18b).

- (18) a. *Ali emruz kār-i na-dāšt, pas kār-i anjām na-dād*  
 Ali today work-IC NEG-have.PST so work-IC finish NEG-give.3SG  
 “Today Ali didn’t have anything to do so he didn’t do anything”

---

2. It is important to consider a much wider set of examples and conduct a more systematic study of the range of constructions and contexts that the object marker appears in (or does not). For a more comprehensive set of examples see Jasbi (2014)

- b. # *Ali emruz kār-i na-dāšt, pas kār-i-ro anjām na-dād*  
 Ali today work-IC NEG-have.PST so work-IC-OM finish NEG-give.3SG

Another prediction is that if the problem in (18b) is truly the clash between the nonexistence claim in the first clause and the existence implication of *rā*, then by changing the first clause to assert that there are tasks or work to do (the set denoted by *kār* is not empty), using the object marker should become acceptable. This is what (19) shows below. (18) and (19) together provide evidence that *rā* encodes an existence implication ( $[[[NP]]] \geq 1$ ).

- (19) *Ali emruz xeyli kār dāšt vali kār-i-ro anjām na-dād*  
 Ali today very work have.PST but work-IC-OM finish NEG-give.3SG  
 “Ali had a lot of work to do but he didn’t do any of them.”

Now let us consider the second aspect of *rā*’s meaning: that the existence implication of *rā* is presuppositional. Example (20) below constructs a context in which the existence of a set of cars that the speaker wants to buy is known to the speaker (Reza) but it is not common ground between the speaker and the addressee (i.e. not presupposed). Given this context, Reza can use an indefinite without the object marker to inform Hasan that he has bought a car (20a) but it is odd to use the object marker (20b).

- (20) [Context: Reza wanted to buy a car and had looked at a couple of models. Hasan did not know about any of this. One day Reza walked in and said:]  
 a. *ye māšhin xarid-am*  
 ID car buy.PST-1SG  
 “I bought a car.”  
 b. # *ye māšhin-o xarid-am*  
 ID car-OM buy.PST-1SG  
 “I bought a car (one of the cars).”

Now we can minimally change the context in (20) such that the set of cars that Reza considered was also known to Hasan. In (21), it is common ground between Reza and Hasan that Reza wants to buy a car and has looked at a couple of options. In other words, a set of cars that Reza can choose from is already presupposed. Now imagine that Reza repeats the same utterances as before.

- (21) [Context: Reza wanted to buy a car and had looked at a couple of models and discussed them with Hasan. They never decided which one is better to buy. One day Reza walked in and said:]  
 a. *ye māšhin xarid-am*  
 ID car buy.PST-1SG  
 “I bought a car (not necessarily one of the discussed cars).”



- b. *ye māšīn-o xarid-am*  
 ID car-OM buy.PST-1SG  
 “I bought a car (one of the discussed cars).”

While both utterances are now acceptable, they do not receive the same interpretation. The absence of the object marker in (21a) suggests that the car Reza bought possibly did not belong to the set he discussed with Hasan. The presence of the object marker in (21b) implies that the car Reza bought was chosen from the same discussed set of cars. This effect is similar to what Enç (1991) reports for Turkish and what Farkas (1994) calls partitive specificity. Examples (20) and (21) show that the existence implication of *rā* is required to be common ground between discourse participants and it is therefore, presuppositional.

Finally, what is the difference between *rā*-marked definite objects (e.g. *māšīn-ō*) and *rā*-marked indefinite objects (e.g. *ye-māšīn-ō*)? I suggest that the answer is uniqueness: *rā*-marked definite objects also carry a uniqueness presupposition ( $[[[NP]] | = 1]$ ) but *rā*-marked indefinites lack a uniqueness implication. Consider Example (22). The context of the example is constructed such that the nominal *muš* “mouse” denotes a unique entity. In such a context, it is completely acceptable if *muš* “mouse” appears without the indefinite determiner *ye* (22a), but odd if the indefinite determiner is present (22b).<sup>3</sup>

(22) [Context: There is a room. Ali goes in. There is a mouse.]

- a. *mush-o mi-bin-e*  
 mouse-OM IPFV-see-3SG  
 “He sees the mouse.”
- b. # *ye mush-o mi-bin-e*  
 ID mouse-OM IPFV-see-3SG  
 “He sees a mouse.”

We can change the context minimally now to have two mice in the room instead of only one. In (23) *mush* “mouse” does not denote a unique entity anymore. In such a context, the absence of the indefinite determiner (23a) makes the utterance unacceptable, but the presence of the indefinite determiner makes a perfectly natural utterance (23b). Examples (22) and (23) suggest that the absence of the indefinite determiner on a *rā*-marked object triggers a uniqueness presupposition.

---

3. The oddness of (22b) is probably due to the anti-uniqueness condition/implication that indefinites generally impose (Heim, 1991).

(23) [Context: There is a room. Ali goes in. There are two mice.]

- a. # *mush-o mi-bin-e*  
 mouse-OM MI-see-3SG  
 “He sees the mouse.”
- b. *ye mush-o mi-bin-e*  
 ID mouse-OM MI-see-3SG  
 “He sees a mouse.”

To summarize, in this section I have shown that the meaning of the object marker *rā* has two components: (1) an existence implication; and (2) that its existence implication is presuppositional. I also showed that the presence/absence of the indefinite determiner *ye* on a *rā*-marked nominal controls the uniqueness presupposition: when *ye* is present the uniqueness presupposition is absent; when it is absent the uniqueness presupposition is present. In the next section I provide a formal analysis of the main findings summarized here.

Before moving to the analysis, I present an unexpected example that follows naturally from the account presented here. For a long time the literature had assumed that *rā* is obligatory on proper names. The explanation was that proper names are definite and *rā* must appear on definite NPs. (24) below shows that *rā* can be optional on a proper name.

- (24) a. [Context: Hasan received a spam-like email from someone named Ali Saburi who claimed is an acquaintance of Reza. He is not sure if Reza knows anyone with this name. He asks Reza:]  
*Ali (e) Saburi mi-šnās-i?*  
 Ali (EZ) Saburi MI-know-2SG  
 “Do you know anyone named Ali Saburi?”
- b. [Context: Ali Saburi is a famous Iranian singer. Hasan wants to know whether Reza knows him. He asks Reza:]  
*Ali (e) Saburi-ro mi-šnās-i?*  
 Ali (EZ) Saburi-OM MI-know-2SG  
 “Do you know Ali Saburi?”

In (24a), the context of the conversation is such that the existence of an entity named ‘Ali Saburi’ cannot be presupposed. The existence of such a person is the main issue in Hasan’s question. However, in (24b) Ali Saburi is a famous singer and Hasan is not asking whether such a person exists or not. He is asking whether Reza knows him. In other words, in (24b) the existence of an entity named ‘Ali Saburi’ is presupposed. This observation follows directly from the account proposed in this section.

#### 4. Formal analysis

In my formal analysis I take the standard approach of formal semantics in translating fragments of natural language into a logical language such as predicate logic (Montague 1973; Heim & Kratzer 1998). I take Persian nominals to be predicates of the type  $\langle e, t \rangle$ . I propose that the indefinite determiner *ye* introduces an existential quantifier, similar to the English indefinite determiner *a* (lexical entry in (25b)). Since Persian has no overt definite article, definite constructions are covertly type-shifted via Partee (1986)'s *iota* operator. Figure 1, from Jasbi (2016), shows the derivation of simple definite and indefinite constructions in the subject position.

Following Coppock & Beaver (2012), I decompose definiteness into two main parts: an existence presupposition and a uniqueness presupposition. I argue that in the object position, the object marker *rā* triggers the existence presupposition. I have shown the lexical entry for *rā* in (25a) using Beaver and Krahmer (2001)'s presupposition operator  $\partial$ . If the sentence lacks the indefinite determiner *ye*, the nominal can be type-shifted via Partee (1986)'s *iota* operator and the derivation can continue, resulting in a definite construction in the object position. Figure 2 shows the complete derivation of a sample definite construction in the object position.

- (25) a.  $r\bar{a} \rightsquigarrow \lambda P [\lambda x [\partial[|P| \geq 1] \wedge P(x)]]$   
 b.  $ye \rightsquigarrow \lambda P \lambda Q [\exists x [P(x) \wedge Q(x)]]$

If the indefinite determiner *ye* is present, it can combine with the *rā*-marked NP and form a generalized quantifier that carries an existential presupposition. Figure 3 shows the complete derivation for a sample object-marked indefinite construction.

#### 5. Discussion

In this chapter, I have argued that the meaning of the Persian object marker *rā* can be best captured by the notion of existential presupposition. If we look closely at the literature on the Persian object marker *rā*, we see that the main components of this proposal had been discussed before. The hypothesis that *rā* carries old or presupposed information forms the main part of Ghomeshi (1996)'s account as well as the information structural accounts of Dabir-Moghaddam (1992) and Dalrymple and Nikolaeva (2011). For Mahootian (1997), *rā* marks some degree of definiteness because even though it appears on indefinite nominals, it is different from ordinary indefinites since it picks out a “delimited class of things in the world”. The hypotheses that *rā* picks a delimited class of objects as well as the hypothesis that it carries old information are both present in the partitive specificity account proposed by Karimi (2003). The main goal of this chapter was to capture the insights

in the previous literature and propose a formal account that can move the debate forward, and inspire new questions on definiteness, specificity, and differential object marking in Persian.

## Abbreviations

1	First Person	NEG	Negation
2	Second Person	NPST	Non-Past Tense
3	Third Person	PL	Plural
EZ	<i>Ezafe</i> Marker	PST	Past Tense
IC	Indefinite Clitic	SG	Singular
ID	Indefinite Determiner	UM	Uniqueness Maker
IPFV	Imperfective Aspect		

## References

- Abbott, Barbara. 2006. Definite and indefinite. *Encyclopedia of language and linguistics* 3. 392–399. <https://doi.org/10.1016/B0-08-044854-2/01089-0>
- Beaver, David & Emiel Kraahmer. 2001. A partial account of presupposition projection. *Journal of Logic, Language and Information* 10(2). 147–182. <https://doi.org/10.1023/A:1008371413822>
- Chierchia, Gennaro & Sally McConnell-Ginet. 1990. *Meaning and grammar: An introduction to semantics*. Cambridge, MA: MIT Press.
- Coppock, Elizabeth & David Beaver. 2012. Weak uniqueness: The only difference between definites and indefinites. In *Proceedings of SALT*, vol. 22.
- Dabir-Moghaddam, Mohammad. 1992. On the (in) dependence of syntax and pragmatics: Evidence from the postposition-*ra* in Persian. In Dieter Stein (ed.), *Cooperating with written texts: The pragmatics and comprehension of written texts*, 549–574. Berlin: Mouton de Gruyter. [https://doi.org/10.1515/9783110881196\\_549](https://doi.org/10.1515/9783110881196_549)
- Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*, vol. 131. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511993473>
- Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22(1). 1–25.
- Farkas, Donka F. 1994. Specificity and scope. In Lea Nash & George Tsoulas (eds.), *Proceedings of Langues et Grammaire* 1. 119–137. Paris: University of Paris 8.
- Farkas, Donka F. & Adrian Brasoveanu. 2013. A typology of specificity. *Revue Roumaine de Linguistique* 58(4). 355–369.
- Fodor, Janet D. & Ivan A. Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5(3). 355–398. <https://doi.org/10.1007/BF00351459>
- Ghomeshi, Jila. 1996. Projection and inflection: A study of Persian phrase structure. Toronto: University of Toronto dissertation.
- Heim, Irene. 1991. Artikel und Definitheit. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der Zeitgenössischen Forschung*, 487–535. Berlin: de Gruyter.
- Heim, Irene. & Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.

- Jasbi, Masoud. 2014. *The semantics of differential object marking in Persian*. Manuscript Stanford: Stanford University.
- Jasbi, Masoud. 2015. The semantics of differential object marking in Persian. Talk given at the 2015 annual meeting of the Linguistics Society of America (LSA 2015). Portland, OR.
- Jasbi, Masoud. 2016. Three types of indefinites in Persian: Simple, complex, and antidefinite. In Mary Moroney, Carol Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Proceedings of Semantics and Linguistic Theory* 26. 244–263. Ithaca, NY: CLC Publications.
- Karimi, Simin. 1990. Obliqueness, specificity, and discourse functions: *Râ* in Persian. *Linguistic Analysis* 20. 139–191.
- Karimi, Simin. 1999. Specificity effect: Evidence from Persian. *Linguistic Review* 16. 125–142. <https://doi.org/10.1515/tlir.1999.16.2.125>
- Karimi, Simin. 2003. On object positions, specificity and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Wiley-Blackwell. <https://doi.org/10.1002/9780470758403.ch5>
- Mahootian, Shahrzad. 1997. *Persian*. Routledge.
- Montague, Richard. 1973. The proper treatment of quantification in ordinary English. In Jaakko Hintikka, Julius Moravcsik & Patrick Suppes (eds.), *Approaches to natural language*, 221–242. Springer. [https://doi.org/10.1007/978-94-010-2506-5\\_10](https://doi.org/10.1007/978-94-010-2506-5_10)
- Nikolaeva, Irina. 2001. Secondary topic as a relation in information structure. *Linguistics* 39(1). 1–50. <https://doi.org/10.1515/ling.2001.006>
- Partee, Barbara. 1986. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jongh & Martin Stokhof (eds.), *Studies in discourse representation theory and the theory of generalized quantifiers*, 115–143. Dordrecht: Foris.
- Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493. <https://doi.org/10.1093/mind/XIV.4.479>
- Stalnaker, Robert C. 1978. Assertion. In Peter Cole (ed.), *Pragmatics* (Syntax and Semantics 9). Academic Press. [https://doi.org/10.1163/9789004368873\\_013](https://doi.org/10.1163/9789004368873_013)
- Strawson, Peter F. 1950. On referring. *Mind* 59(235). 320–344. <https://doi.org/10.1093/mind/LIX.235.320>
- Tonhauser, Judith, David Beaver, Craig Roberts & Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89(1). 66–109. <https://doi.org/10.1353/lan.2013.0001>

## Appendix

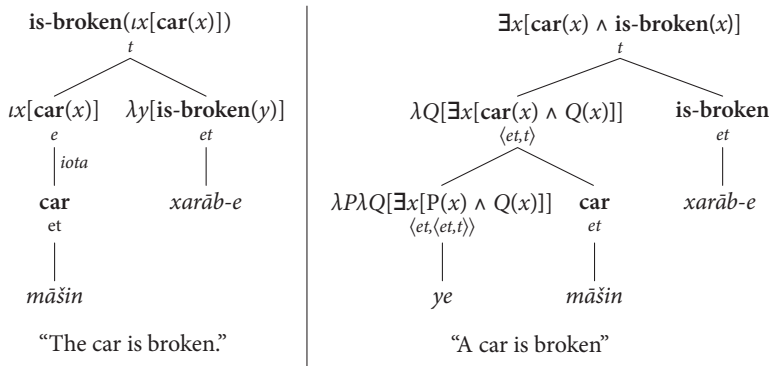


Figure 1. Derivations for sample definite and simple indefinite constructions in Persian

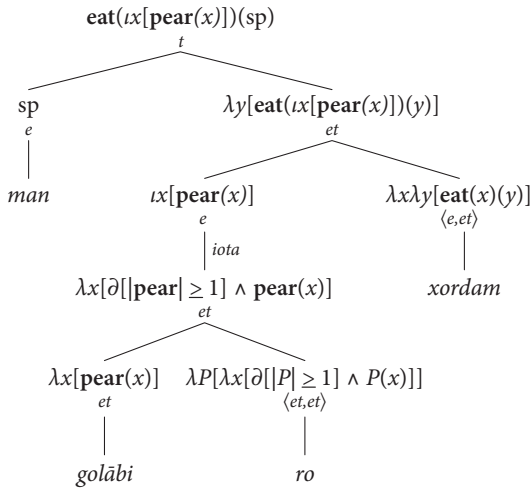


Figure 2. Sample derivation of an object-marked definite in the sentence ‘I ate the pear’

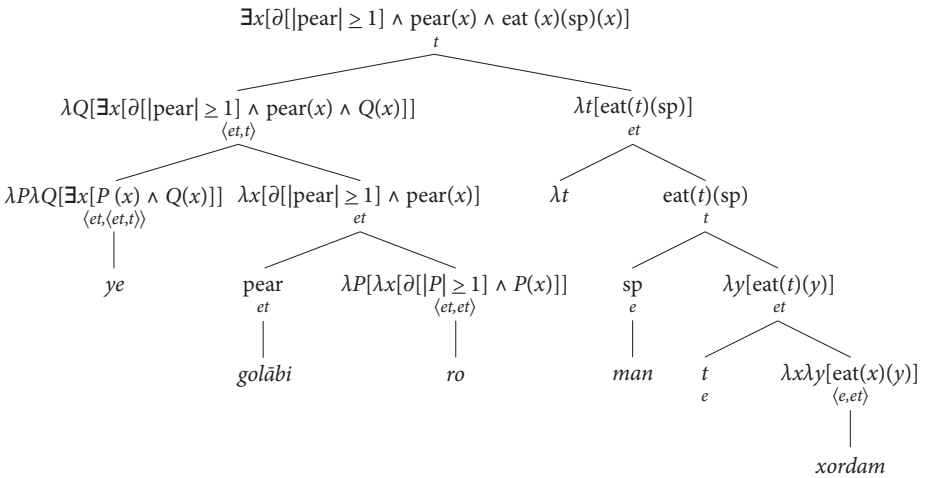


Figure 3. Sample derivation of an object-marked indefinite in the sentence “I ate a certain pear/one of the pears.”



# Topic agreement, experiencer constructions, and the weight of clitics

Thomas Jügel and Pollet Samvelian

Kyoto University / Sorbonne Nouvelle University

It has been claimed that in some Iranian languages like Sorani Kurdish enclitic pronouns shifted to verbal agreement markers via topic agreement, i.e. hanging topics resumed by enclitic pronouns are reanalysed as subjects cross-indexed by agreement markers. In this study, we suggest a bridging context for the reanalysis of topic agreement as verbal agreement by assuming that verbal endings (the inherited agreement markers) and enclitic pronouns represent the same degree or weight of encoding. We further compare the historical findings with the cross-reference patterns found in New Persian experiencer constructions. These constructions show a similar development, and we provide evidence that the relation of experiencer and cross-indexing enclitic pronoun qualifies as agreement.

**Keywords:** enclitic pronouns, agreement, reanalysis, experiencer construction, Middle Iranian, Modern Persian

## 1. Introduction

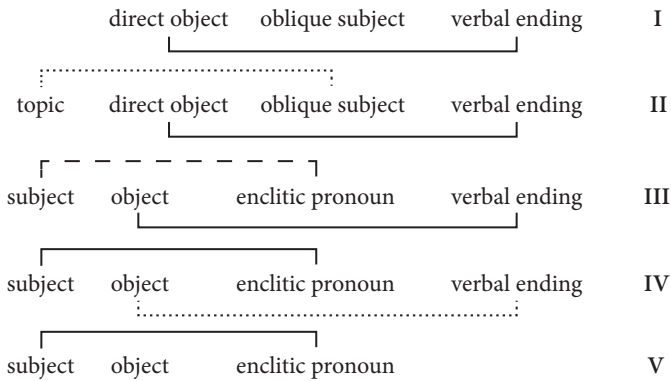
Historically, verbal agreement in Iranian languages is restricted to the nominative subject. The agreement markers are inherited verbal endings that indicate person and number and in combination with further suffixes also tense and mood.

Modern Iranian languages exhibit further sources of agreement markers, viz. auxiliaries, demonstrative pronouns and enclitic pronouns. The most common auxiliary that has turned into a verbal ending is the copula verb ‘be’. In origin a verb stem with verbal endings, this form became first enclitic and eventually a suffix. Demonstrative pronouns were reanalysed as copula expressions, which then continued the same development as the copula verb.

The reanalysis of enclitic pronouns as agreement markers proceeded during the shift of ergative alignment to accusative alignment. In an ergative construction, the subject is encoded by oblique expressions (e.g. enclitic pronouns) and the



object appears in the nominative or direct case agreeing with the verb by means of verbal endings (cf. Figure 1, stage I). It has been suggested earlier (Bynon 1979; Jügel 2009) that the shift proceeded via a hanging topic construction. The subject appeared as a hanging topic and was resumed by the enclitic pronoun (stage II). This topic agreement was reanalysed as subject-verb agreement (stage III) and the previous object-verb agreement was gradually lost (stages IV and V). For New Iranian examples, see Jügel & Samvelian (2016). This scenario is supported by a parallel development in Aramaic (cf. Coghil 2016: 137ff., 230).



**Figure 1.** Historical development of cross-indexing patterns I

In this chapter, we support this scenario with further evidence. We hypothesise that clitic pronouns share relevant features with the inherited agreement markers (the verbal endings), which facilitates their reanalysis as agreement markers (§ 2). Moreover, we think that one can find a similar development in Modern Persian experiencer constructions (§ 3).

## 2. Reanalysis of enclitic pronouns as agreement markers

We assume that clitic pronouns have the same weight (or degree) of encoding as agreement markers. A referent can be expressed by nouns, orthotone pronouns, and clitic pronouns or agreement markers. ‘Expression’ refers here to indexing the referent by morphological means irrespective of the morphosyntactic status of the indexing morpheme. In accusative alignment (verbal forms derived from the present stem), the subject is co-indexed by the agreement marker. In addition, it can be expressed by a noun or pronoun. In an ergative construction (verbal forms derived from the past stem), the subject is not co-indexed. However, Middle Iranian languages permit two expressions that are not possible in accusative alignment,

viz. expression by enclitic pronouns or complete omission (i.e. zero expression). Comparing the expression of the subject referent in accusative and ergative alignment gives the following Figure 2.

accusative construction	NP	+ AGR	high
ergative construction	NP.OBL		
accusative construction	PRON	+ AGR	average
ergative construction	PRON.OBL		
accusative construction	∅	+ AGR	low
ergative construction	PC.OBL		
accusative construction	n/a		null
ergative construction	∅		

Figure 2. Subject expression in accusative and ergative constructions

It is obvious that, if arranged as such, agreement markers and clitic pronouns share the same degree of expression, viz. the lowest degree with an explicit morpheme. This allows us to stipulate the following bridging context: Just as agreement markers co-index an omitted subject, enclitic pronouns co-index an omitted subject (Figure 3).

accusative construction	∅	+ AGR	low
ergative construction	PC.OBL		
<i>reanalysed</i> ↓			
ergative construction	∅	+ PC.AGR	
<i>generalised</i> ↓			
accusative construction	NP	+ AGR	high
ergative construction	NP(.OBL)	+ PC.AGR	
accusative construction	PRON	+ AGR	average
ergative construction	PRON(.OBL)	+ PC.AGR	
accusative construction	∅	+ AGR	low
ergative construction	∅	+ PC.AGR	

Figure 3. Bridging context of enclitic pronouns reanalysed as agreement markers (by Jügel)

This reanalysis is even more likely if oblique case marking of the subject in the ergative construction disappears (which is the case in Middle Persian and Parthian, cf. Jügel 2015: 182) and if, parallel to the structure with omitted subjects, hanging topic constructions appear where the hanging topic can be reanalysed as the conominal of the agreement marker, cf. Examples (1)–(3) with the sequences PC-verb-subject, verb-PC-subject, and subject-verb-PC.

- (1) PC-V-S (AW 101 § 7)  
*u=š guft dādār ohrmazd kū...*  
 and=PC.3SG said.PST creator Ohrmazd that  
 “and he said, the creator Ohrmazd, that...”
- (2) V-PC-S (ZWY 1 § 6)  
*guft=iš ohrmazd ō spitāmān zardušt kū...*  
 said.PST=PC.3SG Ohrmazd to Spitāmān Zardušt that  
 “and he said, Ohrmazd, to Spitāmān Zardušt that...”
- (3) S-V-PC (RPP ar § 4)  
*tahm ud nēw pus ī dōšist wizišt=iš*  
 strong and brave son EZ most\_loved teach.PST=PC.3SG  
*čē=m=iš pursīd*  
 what=PC.1SG=PC.3SG ask.PST  
 “the strong and brave son, the most loved one, he taught what I have asked him to”

The conominals of the enclitic pronoun in Example (1)–(3) are unmarked for case and in Example (2) and (3) the enclitic pronouns appear in the same position where the verbal endings would appear.

## 2.1 Keeping pronouns and agreement markers apart

There is a tendency in typology to abandon the distinction of pronouns and agreement markers. While it is correct that pronouns can become agreement markers and agreement markers pronouns, we have emphasised the necessity of differentiating between the two (Jügel & Samvelian 2016: 408ff.).

In order to illustrate the differences, we have investigated the encoding strategies in a Middle Persian story, the *Book of the Deeds of Ardashīr, Son of Pābag* (KN, edition Čunakova 1987). We wanted to test which degree of encoding is chosen by the author to refer to a participant. Choosing a consistent text was vital, because only then the author has a true choice with respect to the encoding of continued participants. In *Dēnkard Book VI* (Shaked 1979), there appear lists of aphorisms, which are continuously introduced by *u=šān guft ēstēd* (and=PC.3PL say.PST stay.PRS.3SG) “and they have said” or simply “it is said”. These established wordings do not contain continued referents, and the author does not choose to express the subject by an enclitic pronoun or something else. If such examples are included in the overall number of encoding strategies in Middle Persian, they distort the results.

Omission of core arguments is very frequent with the following hierarchy: A > S > P (cf. Table 1). A arguments are more often omitted in accusative alignment (i.e. non-past tenses) than in ergative alignment (i.e. past tenses) which is probably due to the co-indexing agreement marker in the present. This seems to be contradicted by higher omission of S arguments in past tenses. However, S arguments are

**Table 1.** Omission of core arguments in the KN

	Total	Present	Past
A	47% (221)	69% (43)	44% (178)
S	30% (124)	25% (33)	33% (91)
P	15% (79)	18% (23)	14% (56)

co-indexed in non-past as well as past tenses. So the reason for higher omission of S in the past domain is the text layer. Present tense forms occur in dialogues, which are usually short (one or two sentences). The narration is given in past tense and here the narrator tends to continue a topic as subject. Thus subjects in the past domain are frequently omitted, because they continue the topic. Confer the following Example (4) where *Ardaxšir* is the continued A- and S-subject.

- (4) A- and S-subjects (KN 13 § 1–4)  
*Ardaxšir* came back and [Ø] fought the battle with Mihrag and [Ø] killed Mihrag. [Ø] made (Mihrag's) land and property his own. [Ø] sent someone to fight a battle with the Worm. [Ø] called Burzag and Burzādur forth and [Ø] discussed (it) with (them). [Ø] took a lot of money and [Ø] dressed himself like a Khorasanian. [Ø] came with Burzag and Burzādur to the Kulālān fortress and [Ø] said: "..."

Table 1 also shows that the P argument is not affected by the alignment type, which corroborates findings of Jügel (2015: 474), viz. ergativity in Middle Persian only affects the morphological marking but has no effect on syntactic operations.

The following Table 2 displays the frequency of the four degrees of encoding: high = noun phrase, average = orthotone pronoun, low = enclitic pronoun, null = omitted. Beside core arguments, adpositional phrases and possessors (poss., e.g. in an *Ezāfe* construction) are considered as well.

**Table 2.** Encoding degree / participant in the KN

	Total	NP	Ø	Pron	PC
S present	111	46%	30%	24%	
S past	301	50%	44%	7%	
A present	60	8%	72%	20%	
A past	403	29%	44%	5%	22%
P present	105	63%	22%	12%	3%
P past	286	77%	20%	4%	
adpositional phrase	495	87%	<1%	8%	5%
possessor	204	67%	1%	–	–
dative subject	64	39%	42%	8%	14%
indirect object	26	27%	31%	–	–

Pronouns are less preferred strategies of encoding, with subjects tending to be omitted and objects to be expressed by noun phrases.

Non-subjects are preferably encoded by noun phrases with objects showing a considerable amount of omissions (predominantly of inanimate objects). The use of enclitic pronouns to encode objects in the present domain is fairly similar to use of orthotone pronouns encoding objects in the past domain (where the encoding by enclitic pronouns is excluded). However, all enclitic pronouns encode animate objects (3×), while orthotone pronouns evenly encode animate and inanimate objects (9× each).

Orthotone pronouns are preferred for subjects in the present domain, i.e. in dialogues, followed by objects in the present domain. Dative subjects and indirect objects show a fairly even distribution of high encoding and omission.

S arguments tend to be expressed either high or low with a small preference for high. In dialogues (present tense domain) orthotone pronouns are frequent, in narration (past tense domain) omission is more frequent than in dialogues.

A arguments are usually omitted. This is probably because A arguments tend to continue topics rather than expressing new information (cf. du Bois 1987: 826ff.). This tendency is even stronger in dialogues than in narration where pronominal encoding is slightly higher. Encoding by orthotone pronouns in dialogues is fairly even to encoding by enclitic pronouns in narration with very few cases of orthotone pronouns in narration. Thus, in a consistent text, enclitic pronouns appear as often as orthotone pronouns. However, taking orthotone and enclitic pronouns together, pronominal encoding of A is higher in the past tense domain (27%) than in the present tense domain (20%). This is because, in the past tense domain, A arguments appear in ergative constructions and are thus not cross-referenced by an agreement marker. The motivation to express A explicitly is therefore higher than in accusative constructions. In Middle Persian, the narrator has the choice to use the orthotone pronoun (5%) or the enclitic pronoun (22%). The orthotone pronouns usually appear when encoding 1st or 2nd persons, i.e. again in dialogues. Hence, enclitic pronoun marking is the default for A arguments in narration. They are the only case-marked pronouns and therefore the only distinct encoding of the ergative subject.

Comparing A and S arguments, it is apparent that the use of orthotone pronouns is fairly even (5% vs. 7%). If enclitic pronouns were functionally close to agreement markers, their number should be similar to  $\emptyset$  encoding of S arguments (because a  $\emptyset$ -S is still cross-referenced by an agreement marker). However,  $\emptyset$  encoding of S equals the one of A (both 44%). So in contrast to S arguments, A arguments show much more pronominal encoding, but less than S arguments are cross-referenced by agreement markers. For one, this disparity can be explained

by the ergative encoding of A arguments mentioned above. However, we were able to identify two specific contexts where the encoding by enclitic pronouns was the preferred strategy in narration.

a. Continued subject after direct speech

If the continuation of subjects is interrupted by a direct speech, then the A argument is frequently encoded by an enclitic pronoun.

(5) KN 18 § 18–19

**Ardaxšīr** said: “Why did you keep such a good child hidden from me for seven years?” And **he** [PC] held Ohrmazd dear.

b. Continued subject in embedded subclause (including relative clauses)

If the subject is continued as an A argument in a subclause that is inserted in the main clause (often directly following the subject of the main clause), then the A argument is frequently encoded by an enclitic pronoun.

(6) KN 15 § 20

**Ardaxšīr**, when **he** [PC] saw his own child Šābuhr, fell on (his) face and thanked the gods.

Relative clauses also exhibit this pattern, but in contrast to other subclauses they show a much higher amount of omitted A arguments. If we compare the distribution of enclitic pronouns and omitted A arguments in the data of Jügel (2015), the relation is 80% vs. 20% for enclitic pronouns in 849 subclauses in contrast to 52% vs. 48% for enclitic pronouns in 1,022 relative clauses.

## 2.2 Interim results

Subjects tend to be minimally encoded. Agreement markers and enclitic pronouns both represent the lowest degree of encoding. Due to this similarity, enclitic pronouns were reanalysed as agreement markers. The most probable bridging context is a hanging topic construction where the hanging topic is reanalysed as the A argument of the verb and the resuming enclitic pronoun as the index of verbal agreement.

### 3. Experiencer constructions in Modern Persian

Experiencer constructions refer to a class of constructions displaying the following properties:

- a. They refer to a psychological, mental or physical state, implying thus at least an Experiencer (or Beneficiary) argument.
- b. They are formed by a verb and preverbal element, generally a noun or an adjective. The latter conveys the conceptual/lexical meaning of the predicate (e.g. *xande* “laughter”, *qosse* “sorrow”, *sard* “cold”, *hasudi* “jealousy” ...) while the verb is a ‘light verb’ (e.g. *šodan* “become”, *gereftan* “to take”, *zadan* “to hit”, *āmadan* “to come” ...) and has little if no lexical semantic contribution. This has led some studies to consider the sequence formed by the verb and the non-verbal element as a compound verb or a complex predicate (Barjasteh 1983; Rasekhmahand 2010).
- c. The Experiencer is obligatorily encoded by a clitic attached to the preverbal element, and can be cross-referenced by a noun phrase the status of which is disputed. This NP is generally realized in the initial position of the sentence and is comparable to some extent with a hanging topic.
- d. The verb is always in the 3rd person singular.

We think that these constructions represent an interesting parallel to the agreement patterns described in § 2, viz. enclitic pronouns resuming a hanging topic were reanalysed as agreement markers cross-referencing the subject.

Examples (7)–(12) illustrate the Experiencer construction:

- (7) *Xāne-ye edrisihā* (Alizadeh 1991: 23)  
*ādam vahšat=eš mi-gir-ad*  
 human fear=PC.3SG IPFV-take.PRS-3SG  
 “One is afraid.”
- (8) *in pesar be xāhar=eš hasudi=š mi-šod*  
 this boy to sister=PC.3SG jealousy=PC.3SG IPFV-become.PST.3SG  
 “This boy was jealous of his sister.” [lit. “this boy, jealousy of his sister was coming to him”]
- (9) *to be in badbaxt rahm=et ne-mi-ā-d?*  
 you to this miserable pity=PC.2SG NEG-IPFV-come.PRS-3SG  
 “Don’t you have pity for this poor person?” [lit. “you, does pity for this poor person not come to you?”]
- (10) *Hamsāyehā* (Mahmud 1974: 61)  
*az rang=e zard=e bānu delšure=am mi-gir-ad*  
 from color=EZ yellow=EZ Banu worry=PC.1SG IPFV-take.PRS-3SG  
 “I’m worried about Banu’s yellow complexion.”

- (11) *Šabhā-ye Tehrān* (Alizadeh 1999: 108)  
*vali man az kāsigin bad=am ne-mi-ād*  
 but I from kosygin bad=PC.1SG NEG-IPFV-COME.PRS.3SG  
 “But I don’t dislike Kosygin.”
- (12) *Mādarān va doxtarān* (Amrishahi 1998: 84)  
*be ruznāme=i ke šohar=aš be jāy=e xod bar*  
 to newspaper= INDF that husband=PC.3SG to place=PC.3SG self on  
*sandali gozāšte bud māt=aš borde bud*  
 chair put.PP be.PST.3SG dumbstruck=PC.3SG carry.PP be.PST.3SG  
 “[She] stared at the newspaper that her husband had left at his place, on the chair”

These constructions have been highlighted in several studies on Persian and sometimes discussed in detail. Their idiosyncratic properties have led to divergent analyses, as shown by the variety of labels used to refer to them. For some studies, these are merely a subtype of impersonal constructions (Lazard 1957; Vahidian-Kamkar 2004; Karimi 2005; many others). Karimi (2005) calls them “subjectless constructions” and Windfuhr (1979) “indirect (middle) verbs”. Other studies focus on the fact that the predicate is composed of a verb and a preverbal element and label them as “compound verbs of experience” (Barjasteh 1983), “enclitic compound verbs” (Rasekhmahand 2010) or “pronominal complex predicates” (Kazeminejad 2014). Finally, for Sedighi (2010) these are “psychological predicates”.

One of the main topics of controversy is the status of the subject. Although some studies include these constructions in the larger class of impersonal (or subjectless) constructions, some others (Dabir-Moghaddam 1997; Sedighi 2010) claim that the nominal element preceding the verb, the Theme argument, is in fact the subject, hence the regular agreement at the third singular person. Further support for the subjecthood of the nominal element is provided by data from classical Persian (Golchin Arefi 2012), where the Experiencer, realized as an oblique argument introduced by  $=rā^1$ , is not cross-referenced by a clitic pronoun:

- (13) *Golestān* (Saadi, Chapter 8, story 29)  
*ahmaq=rā setāyeš xoš āy-ad*  
 idiot=RA praise pleasant come.PRS-3SG  
 “Idiots are pleased by praise” [*lit.* “To idiots praise comes pleasant”]

1. In Modern Persian, the enclitic  $=rā$  realizes the differential object marking and is obligatory with all definite objects. Historically,  $=rā$  is the phonological reduction of  $rāy$  in Middle Persian, which in turn comes from the Old Persian postposition  $rādi(y)$ , ‘for (the sake of)’, ‘in account of’, ‘concerning’. The enclitic  $=rā$  developed as an indirect object marker in late Middle Persian and Early New Persian and progressively developed into a direct object marker only in the course of several centuries. For a detailed discussion on  $rāy$  in Middle Persian see Jügel (2015: 192–218, 340–342). For a review of the functions and the analysis of  $=rā$  in Modern Persian see Samvelian (2018: 242–256).



Note that this construction has survived in Modern New Persian, but is rarely used:

- (14) *Sag o zemestān-e bolandi* (Parsipour 1973: 127)  
*hālat=i az balāhat=o bohtzadegi dāšt ke binande=rā*  
 attitude=INDF from idiocy=and stupefaction have.PST.3SG that witness=RA  
*xoš ne-mi-āmad*  
 pleasant NEG-IPFV-come.PST.3SG  
 “[He] had such an expression of idiocy and stupefaction that it was not pleasant  
 for the eye-witnesses”

Under this view, the Experiencer argument, when syntactically realized, may be analyzed as a hanging topic in the initial position of the sentence and resumed by the clitic in the clause. A quick survey of the properties of the two constructions shows however noteworthy differences between the hanging topic and the Experiencer NP in the constructions in hand here (see also Sedighi 2010: 113–116).

Like Experiencers, hanging topics occur sentence initially and are resumed by an enclitic pronoun inside the clause:

- (15) *man<sub>i</sub> pedar=am<sub>i</sub> mixād ber-e*  
 PRON.1SG father=PC.1SG want.PRS.IND.3SG go.PRS.SBJV-3SG  
 “My father wants to go.” [*lit.* “I, my father wants to go”]

However, the Experiencer and the hanging topic behave differently in some important respects:

- Experiencers, but not hanging topics, can follow adjuncts, Example (16).
- Hanging topics, unlike Experiencers, cannot occur to the right of the verb, Example (17).
- Experiencers, but not hanging topics, can be the antecedent of a subject-oriented reflexive, Example (18).

- (16) a. Sedighi, (2010: 114 Example 256)  
*diruz tu kelās Ali<sub>i</sub> xāb=eš<sub>i</sub> bord*  
 yesterday in class ali sleep=PC.3SG take.PST.3SG  
 “Yesterday, in the class, Ali fell asleep.”
- b. Sedighi (2010: 114 Example 257)  
 \**diruz tu kelās un zan-e<sub>i</sub> pedar=eš<sub>i</sub> umad*  
 yesterday in class that women-DEF father=PC.3SG come.PST-3SG  
 [Intended meaning] “Yesterday, in the class, that woman, her father came.”
- (17) a. *az in film xoš=am<sub>i</sub> mi-ād man<sub>i</sub>*  
 from this movie pleasant=PC.1SG IPFV-come.PRS.3SG I  
 “Me, I like this movie.”

- b. \**pedar=am<sub>i</sub> fardā mi-ād man<sub>i</sub>*  
 father=PC.1SG tomorrow IPFV-COME.PRS.3SG I  
 [Intended meaning] “MY father will come tomorrow.”
- (18) a. *man<sub>i</sub> xod=am<sub>i</sub> xand=am<sub>i</sub> gereft*  
 I self=PC.1SG laugh=PC.1SG take.PST.3SG  
 “I, myself, laughed.”
- b. \**man<sub>i</sub> xod=am<sub>i</sub> pedar=am<sub>i</sub> raft*  
 I self=PC.1SG father=PC.1SG go.PST.3SG  
 [Intended meaning] “The father of myself left.”

These significant differences between hanging topics and the Experiencer argument not only confirm that the two constructions cannot be assimilated but also bring to light subject-like properties of the Experiencer argument. Crucially, anteceding reflexive pronouns is a typical feature of subjects. Furthermore, the placement properties of the Experiencer are comparable to those of the subject in Persian. Namely, the latter can follow adverbials and occur postverbally. Experiencers display other properties typical of subjects. They can:

- a. be controllees,<sup>2</sup> Example (19).
- b. be omitted in case of clause coordination, if they are coreferent with the subject of the first clause, Example (20).

- (19) Sedighi (2010: 116 Example 261a)  
*soruš<sub>i</sub> ne-mi-xāst xāb=eš<sub>i</sub> be-bar-e*  
 Soroosh NEG=want.PST.3SG sleep=PC.3SG SBJV-carry.PRS-3SG  
 “Soroosh didn’t want to fall asleep”

- (20) Sedighi (2010: 115 Example 258)  
*ki-ā<sub>i</sub> kot na-pušid-an<sub>i</sub> va sard=ešun<sub>i</sub> šod?*  
 who-PL coat NEG-wear.PST-3PL and cold=PC.3PL become.PST.3SG  
 “Who didn’t wear warm clothes and got cold?”

These two properties are typical properties of subjects. Note, on the other hand, that the nominal element preceding the verb, or the Theme, displays none of these typical subject properties.

At this point, we are left with two sets of properties that lead to opposite conclusions:

---

2. Sedighi (2010: 115) mentions further the ability of the Experiencer to be the controller of the null subject in a subordinate clause. Though this is indeed the case, it is not a typical property of subjects, given that direct and prepositional objects may as well be controllers in similar configurations.

- On one hand, the agreement pattern of the Experiencer construction singles out the Theme, or the non-verbal element preceding the verb, as the subject. Diachronic data from Modern Classical Persian points in the same direction.
- On the other hand, the Experiencer, which is clearly not a hanging topic, displays properties cross-linguistically associated to subjects.

The ability to be the trigger of the verbal agreement seems to be the only criterion that seriously undermines the analysis of the Experiencer as the subject. But, what if the enclitic attached to a non-verbal element was in fact an agreement marker? The latter behaves indeed in several respects like an agreement marker, rather than a clitic pronoun. Namely:

1. It is mandatory. This is a typical property of agreement markers, while different types of pronouns constitute one of the options in argument realization.
2. It can be redundant with the constituent it cross-references, that is, the Experiencer. Pronouns, by contrast, are expected to be in complementary distribution.
3. It can refer to an indefinite or negative polarity noun phrase, Example (21). Clitic pronouns, by contrast, refer to definite/anaphoric noun phrases.
4. It cannot alternate with a full pronoun in the *Ezafe* construction, Example (22a). In their pronominal use, on the other hand, clitics can alternate with a full pronoun, Example (22b).

(21) *hičkas<sub>i</sub> xanda=š<sub>i</sub> na-gereft*  
 nobody laugh=PC.3SG NEG-take.PST.3SG  
 “Nobody laughed.”

(22) a. *\*xande=ye to gereft*  
 laughter=EZ you take.PST.3SG  
 [Intended meaning] “You began to laugh”  
 b. *xande=ye to zibā=st*  
 laughter=EZ you beautiful=be.PRS.3SG  
 “Your laughter is beautiful.”

These facts constitute robust arguments for viewing the clitic as an agreement marker in the Experiencer construction. However, such an analysis implies that the Experiencer construction deviates from the canonical pattern of agreement marking in Persian, in that:

- a. The device used to realize agreement is a clitic and not the personal verbal ending, which is the canonical realization of the agreement.
- b. The agreement is not realized on the verb, which is the canonical locus of the realization of the agreement.

- c. The verb still bears agreement features, which implies that, from a merely morphological perspective, agreement is realized twice in the sentence.

The Experiencer construction thus seems to be a transitional construction. The subject features of the Theme can be interpreted in the spirit of Givón's famous quote (1971: 413): "Today's morphology is yesterday's syntax." The Theme was once the nominative subject and the enclitic pronoun encoded a dative experiencer. Due to the higher topicality of the experiencer and the increasing tendency to identify topics as subjects, the Experiencer accumulated more and more subject properties and the Theme lost them. Although Experiencer constructions do not exhibit canonical agreement, the relation of the co-referential noun or pronoun and the enclitic pronoun is best to be seen as subject-verb agreement, while the agreement of the Theme and verb can be considered default marking (see also Dabir-Moghaddam, 1997). In the following subsection we give evidence for the suggested historical development.

#### 4. Historical background of experiencer constructions

Experiencer constructions are already attested in Old Persian. In the following example, the Experiencer appears in genitive/dative case (complete syncretism) in initial position, while the clausal subject is postposed.

- (23) Old Persian (*Darius Behistun Inscription* IV 48–49)  
*awahyā* *paru* *ḡadayāti* *taya*  
 PRON.3SG.GEN/DAT much seem.PRS.SBJV.3SG.ACT REL.NOM.SG  
*manā* *kārtam*  
 PRON.1SG.GEN/DAT do.PP.NOM.SG  
 "it would seem (too) much to him, what I have done" [*lit.* "to him seems much, what I have done"]

This construction is continued to Middle Persian where the oblique case of the experiencer is only visible with enclitic pronouns. The Old Persian example translated would give: *u=š was saḥād čē=m kird* [and=PC.3SG much seem.PRS.3SG.SBJV what=PC.1SG do.PST]. An attested example is the following.

- (24) Middle Persian (*Bundahišn* 21a § 1, cf. Pakzad 2005: 241)  
*mardomān, ka=šān abar mad, ēg=išān ēdōn xwaš*  
 people if=PC.3PL over come.PST.3SG then=PC.3PL so nice  
*sahist hād čiyōn ...*  
 seem.PST SBJV.3SG like  
 "the people, if (the wind) would overcome them, then (it) would seem so nice to them like ..."

Example (24) not only represents an Experiencer construction but also a hanging topic construction. The hanging topic *mardomān* is resumed by enclitic pronouns in the subclause and the main clause.

In the Middle Iranian period, case marking of nouns and orthotone pronouns is reduced and eventually lost. Middle Persian exhibits case marked 1SG pronouns in the oldest texts only (approximately 3rd c. CE). Parthian continues case marked 1SG pronouns, but loses the distinction elsewhere. Bactrian preserves case marked 1SG and 2SG pronouns and seems to restrict the plural marker *-ān* to oblique case. As such, dative subjects can only be clearly indicated by enclitic pronouns for all persons. This is a parallel to ergative subjects.

We assume that Experiencer subjects developed in parallel to ergative subjects. Starting as an ordinary Experiencer (cf. Figure 4, stage I), they acquired subject properties. When resuming a hanging topic (stage II), they could be reinterpreted as agreement markers (stage III). We assume an increasing tendency to express topics as subjects and in combination with the similarity of enclitic pronouns and agreement markers in their weight of encoding (§ 2), it is possible to explain the reanalysis.

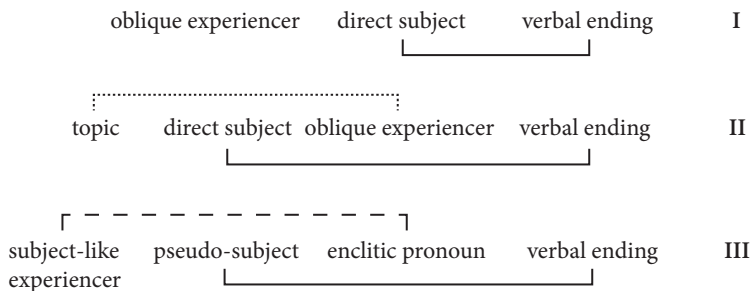


Figure 4. Historical development of cross-indexing patterns II (cf. Figure 1)

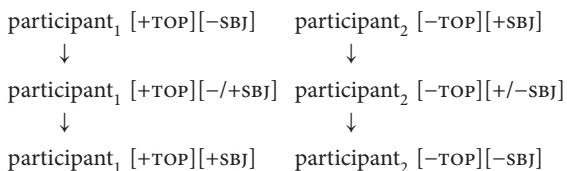
## 5. Summary

We have supported the hypothesis that enclitic pronouns developed into agreement markers via topic agreement by identifying a bridging context, in which agreement markers (verbal endings in accusative constructions) and enclitic pronouns (originally pronouns resuming the hanging topic) show the same weight or degree of encoding. In cases where the agreement marker is not accompanied by a conominal (so-called pro-drop of subject pronouns), they represent the lowest degree of indexing the subject just like enclitic pronouns in an ergative construction. Thus

the enclitic pronouns can be reinterpreted as agreement markers with an omitted conominal and enclitic pronouns resuming a hanging topic can be interpreted as agreement markers with a conominal. In order to understand the differences in the quality of cross-referencing, it is imperative to keep pronouns and agreement markers apart.

The analysis of encoding participants in a Middle Persian text revealed that enclitic pronouns figure very similar to orthotone pronouns, at least where they appear in complementary distribution, i.e. in accusative vs. ergative constructions. Enclitic pronouns do not appear more frequently than orthotone pronouns, so that higher frequency cannot be a way of reanalysing pronouns as agreement markers.

The development of cross-referencing that was identified for transitive constructions can also be stipulated for Experiencer constructions. In both constructions, the topical participant acquires more subject properties and the former subject loses them:



In (prototypical) transitive constructions, participant 1 is the Agent and participant 2 is the Patient. In Experiencer constructions, participant 1 is the Experiencer and participant 2 is the Theme. In contrast to transitive constructions, themes in Experiencer constructions have not lost all subject properties. They preserve nominative case (or rather, they cannot be marked oblique) and they agree with the verb in person and number. However, these features can be considered default forms. In transitive constructions, this is not apparent because the nominative or direct case is unmarked and the verbal ending of the 3SG is a zero ending (i.e. unmarked as well). As such, transitive constructions in languages like Sorani Kurdish can be compared with the situation in New Persian experiencer constructions: the object remains unmarked as does the theme in an Experiencer construction, and the verbal ending is a 3SG in both constructions.

## Funding

This research was funded by a public grant from the French National Research Agency (ANR) as part of the ‘Investissements d’Avenir’ program (Reference: ANR-10-LABX-0083).

## Abbreviations

The glossing follows the Leipzig Glossing Rules with few exceptions indicated below.

ACT	active
AW	Ardā Wirāz Nāmag (Gignoux 1984)
EXP	experiencer
EZ	<i>Ezāfe</i> particle (binding attributes to their antecedent)
KN	Kārnāmag ī Ardaxšīr (Čunakova 1987)
NP	noun phrase
PC	pronominal clitic
Pron	pronoun
PP	past or perfect participle
RPP	Reader in Manichaean Middle Persian and Parthian (Boyce 1975)
RA	marker of the direct object (but cf. footnote 1)
VE	verbal ending
ZWY	Zand ī Wahman Yasn (Cereti 1995)

## References

- Alizadeh, Ghazaleh. 1991. *Xāne-ye edrisihāi*. [Novel.] Tehran: Tus.
- Alizadeh, Ghazaleh. 1999. *Šabhā-ye Tehrān*. [Novel.] Tehran: Tus.
- Amrishahi, Mahshid. 1998. *Mādarān va doxtarān*. [Novel.] Stockholm: Ferdosi.
- Barjasteh, Darab. 1983. Morphology, syntax and semantics of Persian compound verbs: A lexical approach. University of Illinois dissertation.
- Boyce, Mary. 1975. *A reader in Manichaean Middle Persian and Parthian*. Téhéran: E. J. Brill.
- Bynon, Theodora. 1979. The ergative construction in Kurdish. *Bulletin of the School of Oriental and African Studies* 42. 211–224. <https://doi.org/10.1017/S0041977X0014577X>
- Cereti, Carlo G. 1995. *The Zand ī Wahman Yasn – A Zoroastrian apocalypse*. Roma: Istituto Italiano per il Medio ed Estremo Oriente.
- Coghill, Eleanor. 2016. *The rise & fall of ergativity in Aramaic – Cycles of alignment change*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198723806.001.0001>
- Čunakova, Olga M. 1987. *Kniga dejanij Ardašira syna Papaka*. Moskau: NAUK.
- Dabir-Moghaddam, Mohammad. 1997. Compound verbs in Persian. *Studies in the Linguistic Sciences* 27. 25–59.
- Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855. <https://doi.org/10.2307/415719>
- Gignoux, Philippe. 1984. *Le livre d'Ardā Virāz – translitération, transcription et traduction du texte pehlevi*. Paris: Editions Recherche sur les Civilisations.
- Givón, Talmy. 1971. Historical syntax and synchronic morphology: An archaeologist's field trip. *Papers from the 7th Regional Meeting Chicago Linguistic Society, April 16–18, 1971*, 394–415. Chicago: Chicago Linguistic Society.
- Golchin Arefi, Mandeh. 2012. Barresi-e sāxt-e qeyr-e šaxsi dar zabān-e fārsi. *Vijenāme-ye Farhangestan* 7. 162–182.

- Jügel, Thomas. 2009. Ergative Remnants in Sorani Kurdish?. *Orientalia Suecana* 58. 142–158.
- Jügel, Thomas. 2015. *Die Entwicklung der Ergativkonstruktion im Alt- und Mitteliranischen – eine korpusbasierte Untersuchung zu Kasus, Kongruenz und Satzbau*. Wiesbaden: Harrassowitz. <https://doi.org/10.2307/j.ctvc770qq>
- Jügel, Thomas & Pollet Samvelian. 2016. Les «pronoms» clitiques dans les langues iraniennes occidentales. *Bulletin de la Société de Linguistique de Paris* 111(1). 391–432.
- Karimi, Simin. 2005. *A Minimalist approach to scrambling: Evidence from Persian*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110199796>
- Kazeminejad, Ghazaleh. 2014. *Pronominal complex predicates in Colloquial Persian*. Lexington, KY: University of Kentucky.
- Lazard, Gilbert. 1957. *Grammaire du persan contemporain*. Paris: Klincksieck.
- Mahmud, Ahmad. 1974. *Hamsāyehā*. [Novel.] Tehran: Amir Kabir.
- Pakzad, Fazlollah (ed.). 2005. *Bundahišn – Zoroastrische Kosmogonie und Kosmologie I: Kritische Edition*. Tehran: Centre for the Great Islamic Encyclopaedia.
- Parsipour, Shahrnush. 1973. *Sag o zemestān-e bolandi*. [Novel.] Tehran: Amir Kabir.
- Rasekhmahand, Mohammad. 2010. Persian clitics adjacent to verbs. *Journal of Researches in Linguistics Studies* 2(2). 75–85.
- Samvelian, Pollet. 2018. Specific features of Persian syntax: The Ezafe construction, differential object marking and complex predicates. In Anousha Sedighi & Pouneh Shabani-Jadidi (eds.), *Oxford handbook of Persian linguistics*, 226–269. Oxford: Oxford University Press.
- Sedighi, Anousha. 2010. *Agreement restrictions in Persian*. Leiden: Leiden University Press.
- Shaked, Shaul. 1979. *The wisdom of the Sasanian sages (Dēnkard VI)*. Boulder, CO: Westview Press.
- Vahidian-Kamkar, Taghi. 2004. Fe'lhā-ye yek šenāse. *Nāme-ye Farhangestān-e zabān* 2/6. 29–37.
- Windfuhr, Gernot L. 1979. *Persian grammar: History and state of its study*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110800425>





# Another look at Persian *rā*

## A single formal analysis of a multi-functional morpheme

Simin Karimi and Ryan Walter Smith  
University of Arizona

The morpheme *-rā* has typically been analyzed as an instance of differential object marking, appearing on direct objects that are definite or in some sense specific. In this chapter, we discuss several cases in which *-rā* may appear on DPs that are not direct objects. Building on insights from dependent case theory (Marantz 1991; Baker & Vinokurova 2010, Preminger 2011, 2014, Kornfilt & Preminger 2014), we develop an analysis of *-rā*, according to which it is the realization of accusative case, treated as a dependent case assigned in syntax, and a specificity feature. In addition to accounting for the facts discussed in the context of Modern Persian, we show how an extension of our analysis can account for the distribution of *-rā* in Classical Modern Persian.

**Keywords:** *-rā*, specificity, accusative case, dependent case theory

### 1. Introduction

Cross-linguistically, there are two classes of objects: overtly marked, known as Differential Object Marking (DOM), and unmarked. DOM may take the form of a case marker (e.g., Hindi, Turkish, Hebrew), an adposition (e.g., Spanish), agreement (e.g., Swahili, Senaya), or clitic-doubling (e.g., Macedonian, Catalan). Universally, common factors distinguishing objects are definiteness, specificity, and animacy. In general, objects ‘high’ on the relevant scale (e.g., more definite) are marked. One well-known instance of DOM is found in Hindi, where objects are differentiated based (mainly) on specificity: with *-ko* (which is also the canonical dative case marker) when they are specific (Bhatt 2007).

In Persian, the morpheme *-rā* has been typically treated as a differential object marker which appears on specific direct objects (Browne 1970; Karimi 1990), or definite objects (Mahootian 1997; Ghomeshi 1996; among others). There are,

however, several cases in which the morpheme *-rā* appears on DPs other than the direct object, including (i) subjects raised out of an embedded clause, (ii) DPs corresponding to a clitic inside an object, (iii) DPs corresponding to a clitic object of a preposition, and (iv) nominal adverbials. In some of these cases, the predicate is in fact unergative, rather than transitive, making it problematic to analyze every case of a *rā*-marked DP as DOM *per se*. This leads to the following questions: what is the real function of *-rā*? What do DPs marked with *-rā* have in common?

The goal of this article is to propose a formal analysis that explains the distribution of the morpheme *-rā* in a natural and explanatory fashion. To achieve this goal, we analyze DP+*rā* within the framework of *dependent case theory*, according to which dependent cases, such as accusative, are assigned to a DP c-commanded by another DP in the same domain. Based on the data, we motivate a new analysis of *-rā* which indicates that this element marks specific DPs that have been valued for dependent case.

The analysis in this work builds on work by Preminger (2011, 2014) and Kornfilt & Preminger (2014), who argue, on the basis of the Turkic language Sakha, that nominative (as well as absolutive) is simply the morphological form afforded to noun phrases whose case features have not been valued in the course of the derivation. This means that subject DPs are not valued for case. The theory adopted in this work predicts that subjects of embedded clauses raised into the matrix clause may only appear with *-rā* if the matrix clause contains a DP that c-commands the raised DP. We show that this prediction is borne out.

The article is organized as follows. The relevant data are introduced in § 2, followed by a discussion of the theoretical background in § 3. The proposed analysis appears in § 4. Some predictions based on the proposed analysis in this article are discussed and shown to be borne out in § 5. Section 6 introduces data from Classical Modern Persian which are still employed in formal literary contexts. We show in that section that an extension of our proposal accounts for these archaic forms as well. Concluding remarks appear in § 7.

## 2. Data

It is well-known that specific/definite objects, but not nonspecific ones, are marked in Persian. Furthermore, *-rā* is obligatory if the DP is specific/definite.

- (1) *Kimea be man ketāb dād*  
 Kimea to me book gave  
 ‘Kimea gave me (a) book/books.’

- (2) *Kimea in ketāb \*(ro) be man dād*  
 Kimea this book *rā* to me gave  
 “Kimea gave me this book.”

Neither matrix nor embedded subjects may be marked by *-rā*,<sup>1</sup> nor may objects of prepositions be marked with it.

- (3) *Kimea(\*-ro) ketāb xund*  
 Kimea-*rā* book read  
 “Kimea read books.”
- (4) *man fekr mi-kon-am ke Ali(\*-ro) barande mi-sh-e*  
 I thought ASP-do-1SG that Ali-*rā* winner ASP-become-3SG  
 “I know Ali will win (become a winner).”
- (5) *Kimea be Parviz(\*-ro) goft*  
 Kimea to Parviz-*rā* said  
 “Kimea told Parviz.”

At first glance, it seems that the distribution of *-rā* is straightforward: it appears only on definite/specific direct objects. However, there are several cases that complicate this seemingly simple state of affairs. First of all, embedded subjects may be marked by *-rā* if raised into the higher clause. In (6), the raised subject has moved into the main clause.

- (6) *Ali-ro fekr mi-kon-am barande be-sh-e*  
 Ali-*rā* thought ASP-do-1SG winner SBJV-become-3SG  
 “As for Ali, I think he wins”

Topicalized DPs corresponding to the pronominal object clitic of a preposition are also marked by *-rā*.

- (7) *man [Pari-ro]<sub>i</sub> bā-hāsh<sub>i</sub> harf zad-am*  
 I Pari-*rā* with-her talk hit-1SG  
 “As for Pari, I talked with her.”

Similarly, DPs corresponding to clitics inside an object are marked by *-rā* as well.

---

1. Abbreviations: ASP = aspect, SG = singular, PL = plural, NEG = negation, SBJV = subjunctive, INDF = indefinite, EZ = *Ezāfe*, a particle that links a nominal head to its complements/modifiers. See Samiiān (1983, 1994), Ghomeshi (1997), Karimi & Brame (2013), Kahnemuyipour (2014) for various approaches to this element in Persian.

- (8) *man [māshin-ro]<sub>i</sub> dar- esh<sub>i</sub>-ro bast-am*  
 I car - rā door- its-rā closed-1SG  
 ‘As for the car, I closed its door.’ (Karimi 1989)
- (9) a. *pro māmān-e Ali-ro did-am*  
 mother-EZ Ali-rā saw-1SG  
 ‘I saw Ali’s mom.’  
 b. *pro [Ali-ro]<sub>I</sub> māmān-esh<sub>i</sub>-ro did-am.*  
*Ali-ro mother -his-rā saw-1SG*  
 ‘As for Ali, I saw his mom’

Note, however, that the same pattern does not hold when the topicalized DP corresponds to a clitic pronominal inside a subject.

- (10) a. *xāhar -e Sahar (\*-ro) mi-ā-d.*  
 sister EZ Sahar ASP-come-3.SG  
 ‘Sahar’s sister comes.’  
 b. *Sahar (\*-ro) xāhar-esh mi-ā-d,*  
 Sahar -rā sister-her ASP-come-3SG  
 ‘As for Sahar, her sister will come.’

Finally, nominal adverbs, such as the temporal adverbs *fardā* ‘tomorrow’ and *shab-e pish* ‘last night’, as well as directional adverbs like *in rāh* ‘this way’, may be marked by *-rā*, even in the absence of a transitive verb (Karimi 1997).

- (11) a. *man fardā -ro tu xune mi-mun-am*  
 I tomorrow -rā in house ASP-stay-1SG  
 ‘As for tomorrow, I will stay at home.’  
 b. *pro shab-e pish-o aslan na- xābid-am*  
 night-EZ last-rā at all NEG- slept-1SG  
 ‘As for last night, I didn’t sleep at all.’
- (12) *mā in rāh-ro bā ham raft-im*  
 we this way-rā with each other went-1PL  
 ‘We have gone this way with each other.’

### 3. Theoretical background

In this section, we review a few theories with respect to case, and introduce the one we adopt to account for the various data we introduced in the previous section.

Within the Minimalist Program (Chomsky 1995) and subsequent work (Chomsky 2001), Case is seen as a semantically uninterpretable feature on nominals, thereby requiring ‘deletion’ before the semantic interface (LF).

- (13) “Structural Case is not a feature of the probes (T, *v*), but it is assigned a value under agreement. The value assigned depends on the probe: nominative for T, accusative for *v*.” (Chomsky 2001: 6)

There are, however, other approaches to case assignment which consider accusative case as a DEPENDENT case, and do NOT take unmarked cases like nominative to be positively specified. Marantz’s (1991) *disjunctive Case hierarchy* is a prominent example. The portion of Marantz’s proposal that is relevant to our discussion appears in (14).

- (14) Marantz’s Disjunctive Case hierarchy
- i. **Dependent Case:** assigned to a position governed by V+I when a distinct position is governed by V+I is a) not marked and b) distinct from the chain being assigned dependent case (ergative = dependent case assigned up to subject; accusative = dependent case assigned down to object)
  - ii. **Unmarked Case:** assigned when a DP appears embedded in a certain structural position (genitive in NPs, nominative in Spec-IP/TP).

For Marantz, case assignment is a post-syntactic process that applies to the output of syntactic operations. Preminger (2011, 2014), on the other hand, gives the same case assignment algorithm a purely syntactic implementation. In this implementation, dependent case is assigned to DPs that are either c-commanded by another DP within the same domain (accusative case) or c-command another DP (ergative case). Nominative and absolutive cases are simply the morphological form given to DPs that are unvalued for case.

Discussing case within the same school of thought, Baker and Vinokurova (2010), Kornfilt & Preminger (2014) as well as Baker (2017) show that accusative in Sakha, a Turkic language, can only be analyzed as *dependent case*. Consider the example in (15).

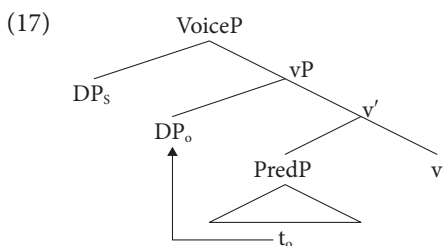
- (15) a. *Min sarsyn ehigi-(\*ni) kel-iex-xit dien ihit-ti-m*  
 I tomorrow you-(\*ACC) come-FUT-2SG that hear-PST-1SG  
 “I heard that tomorrow you will come.”
- b. *Min ehigi-ni [bügün kyaj-yax-xyt dien erem-mit-im.*  
 I you-ACC today win-FUT-2SG that hope-PTCL-1SG  
 “I hoped that you would win today.” (Baker 2017)

(15a) shows that a subject properly contained in an embedded clause cannot receive accusative case in Sakha. When the subject moves to the edge of the embedded clause, as in (15b), it can receive accusative case due to case competition with the subject DP.

In this work, we follow the works cited above in maintaining that accusative case is assigned in a way suggested in (16).

- (16) Case valuation
- Accusative case is a dependent case assigned to a DP c-commanded by another DP (overt or covert) within the same domain.
  - Prepositions assign lexical/oblique case.
  - Nominative case is the absence of case.

Following Karimi (2005), we assume that both specific and nonspecific objects are base-generated inside the PredP (=VP). The specific object moves into a higher position, possibly the Specifier of vP, to escape the novelty domain (Heim 1982; Diesing 1992; Holmberg & Nikanne 2002).



Finally, we suggest a post-syntactic rule of *rā*-marking, as in (18):

- (18) Post-syntactic *rā*-Marking  
 DP<sub>Specific+Accusative</sub> is marked by *-rā* at the morphological interface post-syntactically.

As for the clausal word order in this language, we adopt a slightly revised version of the structure proposed by Karimi (2005), reproduced here in (19).

- (19) [CP[TopP[FP[TP [T'[VoiceP [vP [v' [XP [X' ] v ]]]]]]]]]

Following Karimi (2005), we assume two topic positions in this language: the Specifiers of TopP and TP.

There is one final issue we need to address. Our definition of specificity is based on Enç (1991).<sup>2</sup> She defines specificity in terms of 'strong' and 'weak' antecedents. A summary of our Enç-based definition of specificity is stated in (20).

- (20) Specificity
- A 'definite' DP requires a 'strong antecedent' based on an 'identity' relation between this type of DP and its previously established discourse referent. Therefore, definite DPs are always specific.

2. See also Jasbi (this volume) for a discussion of the variety of specificity relevant for *rā*-marked nominals in Persian.

- An ‘indefinite’ DP is specific if it denotes an inclusion relation to previously established discourse, representing a ‘weak antecedent’.
- A ‘nonspecific’ DP lacks an antecedent in the discourse altogether.

In the next section, we examine the data discussed in § 2 within the framework adopted in this section.

#### 4. Analysis

We start the analysis of the data with the most obvious cases, namely specific direct objects. The example in (2) is repeated here in (21). The object, still inside the vP, is valued for accusative case via case competition with the subject in the specifier of VoiceP.

- (21) *Kimea* [<sub>vP</sub> *in ketāb* \*(-ro) [<sub>PredP</sub> *be man dād*]]  
 Kimea this book *rā* to me gave  
 “Kimea gave me this book”

This analysis is extended to those cases with double DP+*rā*, as in (8), repeated in (22).

- (22) *man* [*māshin-ro*]<sub>i</sub> *dar -esh*<sub>i</sub>-ro *bast-am*  
 I car – *rā* door-its-*rā* close-1SG  
 “As for the car, I closed its door.” (Karimi 1989)

*māshin-ro* corresponds to the clitic inside the object. We suggest that this DP is merged inside the vP, possibly in the Specifier of that phrase, as in (23), and is assigned accusative case, again due to being c-commanded by the DP in SpecVoiceP. The lower DP *dar-esh* is also assigned accusative by virtue of being c-commanded by *māshin*.

- (23) [<sub>VoiceP</sub> *man* [<sub>vP</sub> [*māshin-ro*]<sub>i</sub> [*dar -esh*<sub>i</sub>-ro [*bast-am* ]]]]]

As for the object of the preposition, the statement in (16b) correctly predicts that it cannot be marked by *-rā* since it is embedded inside PP. The example in (4), repeated as (24) exemplifies this fact:

- (24) *Kimea* [<sub>PP</sub> *be Parviz* \*(-ro)] *goft*  
 Kimea to Parviz -*rā* said  
 “Kimea told Parviz.”

The DP+*rā* in (7), repeated in (25), corresponds to a clitic object inside PP. We suggest that this DP, similar to the one in (23), is merged in the Specifier of vP, is assigned accusative by virtue of being c-commanded by the subject DP in SpecVoiceP, and is marked by *-rā* post-syntactically.



- (25) [<sub>VoiceP</sub> *man* [<sub>vP</sub> *Pari-ro*]<sub>i</sub> [*bā-hāsh*]<sub>i</sub> *harf zad-am*]  
 I Pari-rā with-her talk hit-1SG  
 ‘As for Pari, I talked with her.’

This DP, similar to other DPs marked by *-rā*, may move into a higher position, such as the Specifier of TP, preceding the subject, as in (26).

- (26) [<sub>TP</sub> *Pari-ro*]<sub>i</sub> [<sub>VoiceP</sub> *man* [<sub>vP</sub> *t*]<sub>i</sub> [<sub>PrepP</sub> [*bā-hāsh*]<sub>i</sub>] *harf zad-am*]]

Next, consider the case of non-object DPs in an intransitive construction, as in (12), repeated in (27).

- (27) *mā* [<sub>vP</sub> [<sub>in</sub> *rāh*]<sub>i</sub>]-*ro* *bā ham raft-im*  
 we this way-rā with each other went-1PL  
 ‘We have gone this way with each other.’

The appearance of *-rā* on these DPs is a natural consequence of case competition: they are assigned a dependent case due to the fact that they do not appear as complement to an assigner of oblique or lexical case (such as a preposition), and they are c-commanded by a DP in the same domain, in this case the subject DP *mā*.

Next we discuss the case of nominal adverbials. Cinque (1999) suggests a sequence of High and Low adverbials to appear either at the edge or inside the verb phrase. Karimi (2005) has applied Cinque’s model to Persian adverbial phrases. Based on those proposals, we assume that adverbs are either adjoined to vP (high adverbials) or inside it (low adverbials). Thus they may be assigned accusative case, due to being c-commanded by the DP in SpecVoiceP. This analysis is borne out by the data in (11), restated in (28).

- (28) a. *man* [<sub>vP</sub> *farda*]-*ro* *tu xune mi-mun-am*  
 I tomorrow-rā in house ASP-stay-1SG  
 ‘As for tomorrow, I will stay at home.’  
 b. *pro* [<sub>vP</sub> *shab-e pish-o aslan na-xābid-am*]  
 night-EZ last-rā at all NEG- slept-1.SG  
 ‘As for last night, I didn’t sleep at all.’

Finally, the example in (3), restated in (29), shows that the subject DP cannot be marked by *-rā*. This follows from the fact that Persian is a language in which dependent case is assigned downward to a c-commanded DP in a case competition relationship, not the c-commanding DP. As such, subject DPs are not assigned case at all, and *-rā*, being possible only when a DP is both dependent and specific, is correctly predicted not to appear on subjects.

- (29) [<sub>VoiceP</sub> *Kimea*-(*\*ro*) [<sub>vP</sub> *ketāb xund* ]]  
 Kimea-*rā* book read  
 “Kimea read books.”

Note that the DP corresponding to the clitic pronoun inside the subject in (10), repeated in (30b), cannot be marked either.

- (30) a. [<sub>VoiceP</sub> *xāhar-e Sahar* (*\*-ro*) [<sub>vP</sub> *mi- ā -d* ]]  
 sister-EZ Sahar -*rā* ASP-come-3SG  
 “Sahar’s sister comes.”  
 b. *Sahar<sub>i</sub>* (*\*-ro*) *xāhar-esh<sub>i</sub> mi- ā -d*  
 Sahar -*rā* sister-her ASP-come-3SG  
 “As for Sahar, her sister will come.”

This falls out directly from the analysis, as the DP corresponding to the clitic pronoun is not *c*-commanded by any other DP within the same domain. In these cases too, the subject is not assigned any case, and the absence of *-rā* in this context is therefore predicted.<sup>3</sup>

In (6), restated in (31), the embedded subject appears in the main clause and is marked by *-rā*. Note that unlike the data from Sakha where the embedded subject appears at the edge of its own clause (cf. (15)), the subject in Persian may move all the way into the higher clause. We suggest that the embedded subject has moved cyclically through the Specifier of various phases, including the matrix *vP*, and is assigned accusative case in that position through case competition with the matrix subject. The assignment of accusative to the embedded subject is possible on this approach due to the fact that subjects are not assigned case in SpecTP.

- (31) *pro*<sub>[vP]</sub> [*Ali-ro*]<sub>i</sub> *fekr mi-kon-am [barande be-sh-e]*  
 Ali-*rā* thought ASP-do-1.SG winner SBJV-become-3SG  
 “As for Ali, I think he wins.”

Now consider the examples in (32) once again. As discussed before, neither the subject nor the DP corresponding to the clitic pronoun inside the subject may be marked by *-rā*.

---

3. This sentence does raise another issue, however: because the subject DP *xāharesh* is *c*-commanded by another DP, *Sahar*, in the same domain, one might expect that the *c*-commanded DP should be able to be marked by *-rā*, contrary to fact. One possible explanation for this is that the higher DP is in fact not in the same domain as the subject DP in this example, but in a higher Specifier position within the left periphery of the clause. If there is a phasal boundary between the two DPs, then they will not be in the same domain and thus will not be able to enter into case competition with one another.

- (32) a. [<sub>VoiceP</sub> *xāhar-e Sahar* (\*-ro) [<sub>vP</sub> *mi-ā-d*]]  
 sister-EZ Sahar -rā ASP-come-3SG  
 “Sahar’s sister comes.”
- b. *Sahar<sub>i</sub>* (\*-ro) *xāhar-esh<sub>i</sub> mi-ā-d*  
 Sahar -rā sister-her ASP-come-3SG  
 “As for Sahar, her sister will come.”

However, if the topicalized DP appears in the matrix clause, it can be marked by *-rā*, as in (33).

- (33) *Sahar<sub>i</sub>-ro man fekr mi-kon-am xāhar-esh<sub>i</sub> mi-ā-d*  
 Sahar-rā I thought ASP-do-1SG sister-her ASP-come-3SG  
 “As for Sahar, I think her sister will come.”

This is not surprising if the DP moves through the matrix vP, and is assigned accusative case on its way to the topic position in the matrix clause.

An important issue to discuss at this point concerns the status of the raised subject. As the example in (34) shows, marking of an embedded subject that has raised into the matrix clause appears to be optional.

- (34) *Kimea(-ro) man fekr mi-kon-am bā mā bi-ā-d*  
 Kimea(-rā) I thought ASP-do-1SG with us SBJV-come-3SG  
 “As for Kimea, I think she will come with us.”

We suggest that the unmarked version of the embedded subject is merged directly in the specifier of TopP in (34). Since the sentential topic position is not c-commanded by another DP, as in (19), repeated here in (35), the base generated DP cannot be assigned accusative case.

- (35) [CP [TopP [FP [TP [T' [VoiceP [vP [v' [XP [X' ] ] v ]]]]]]]]]

One piece of evidence in favor of a movement analysis in the case of (33) and (34), where the embedded subject is marked by *-rā* in the matrix clause, comes from sentences like (36) and (37), where the raised subject appears in an intermediate position between the matrix subject and predicate.

- (36) *man<sub>i</sub> Ali-ro fekr mi-kon-am barande be-sh -e*  
 Ali-rā thought ASP-do-1SG winner SBJV-become-3SG  
 “As for Ali, I think he wins”

- (37) \**man [vP [Ali]<sub>i</sub> fekr mi-kon-am [ e<sub>i</sub> barande be-sh-e ]]*

While the raised DP + *rā* may appear in this intermediate position below the matrix subject, the unmarked DP in (37) cannot. This can be explained if the *rā*-marked

embedded subject raises from the embedded clause to a lower position in the matrix clause, such as the specifier of vP, where it is c-commanded by the matrix subject. The ungrammaticality of (37) shows that this DP cannot be unmarked.

## 5. Prediction

The analysis we have motivated thus far predicts that a subject DP raised into a clause with an impersonal predicate may not be marked by *-rā*. This is so because such clauses do not possess DP arguments. Thus due to the absence of other DPs in the matrix clause, case competition does not exist, and therefore dependent case assignment should not be possible. The sentence in (38)–(40) demonstrate that this prediction is borne out: the presence of *-rā* on the raised subject is ungrammatical in each case.<sup>4</sup>

- (38) *Ali (\*-ro) ghat'i-e barande mi-sh-e*  
 Ali -rā certain-is winner ASP-become-3SG  
 “As for Ali, it is certain that he wins.”
- (39) *Ali (\*-ro) vāzeh-e barande mi-sh-e*  
 Ali -rā obvious-is winner ASP-become-3SG  
 “As for Ali, it is obvious that he wins.”
- (40) *Ali(\*-ro) be nazar mi-yā-d ketāb-ro xunde bash-e*  
 Ali-rā to view ASP-come-3SG book-rā read be.SBJV-3SG  
 “As for Ali, it seems (he) has read the book.”

At this juncture, it should be noted that there is an alternative analysis for these and the other facts concerning the possibility of *-rā*-marking on embedded subjects raised into the matrix clause; namely, the sorts of predicates that permit *-rā* on raised subjects are all predicates with external arguments, while the predicates in (38)–(40) are not. As such, one might instead propose an analysis in terms of structural Case assignment in which a head, such a *v* or Voice, introduces the external argument and assigns Accusative Case. This head is present with predicates like *fekr kardan* “think”, but not with predicates like *be nazar āmadan* “seem”, and therefore it will be plausible to assign Accusative Case to raised subjects if a matrix clause contains the former but not if it contains the latter.

---

4. We thank several native speakers who provided us with judgments regarding the data in (38)–(41).

At a conceptual level, a structural Case approach along these lines is already problematic, given the analysis of nominative as the absence of case, as DPs must receive Case on such theories in order to be licensed. However, it is also possible to distinguish the structural Case and dependent case theories empirically, making use of unaccusative predicates that select a DP as well as a CP argument. Since these predicates do not have external arguments, the functional head introducing the external argument and capable of valuing a DP for Accusative Case is not present, and we should therefore expect *-rā*-marking to be ungrammatical on subjects raised from the embedded clause. On the other hand, a dependent case approach predicts that *rā* SHOULD be grammatical on a raised subject if another DP is present to c-command it, whether the predicate is unaccusative or otherwise. This is because the conditions for dependent case assignment are met whenever a DP c-commands another DP in the same domain.

Now consider the sentence in (41).

- (41) <sup>?</sup>*man Ali-ro xoshhāl shod -am mi-ā-d*  
 I Ali-*rā* happy became -1SG ASP-come-3SG  
 “As for Ali, I became happy that he is coming.”

Speakers vary somewhat in how they judge (41): some speakers judge it as fully grammatical, while others judge it as somewhat marginal. However, speakers agree that it is more acceptable than (38)–(40) with *rā*-marking on the raised subject. Given that speakers generally rate these sentences as acceptable, we take this to be evidence in favor of the dependent case analysis we have developed in this chapter, and against a structural Case analysis.

The next section examines some non-objective DP+*rā* cases in Classical Modern Persian (CMP), and shows that the proposal can be extended to the analyses of these cases as well.

## 6. Classical Modern Persian

Classical Modern Persian (CMP) offers interesting data marking DPs with the morpheme *-rā* in contexts where the DP is not a direct object of the predicate. We show in this section that the theoretical framework adopted in this article can account for all those archaic cases.

In Old Persian, *-rā* appears as *rādi* marking a cause with the meaning ‘for the sake of’. The same interpretation holds for *rāy*, the reflex of *rādi* in Middle Persian. According to Brunner (1977), Middle Persian *rāy* served other functions as well. It appeared as an illustration of purpose, reference, beneficiary or indirect object (Karimi 1990).

In early Classical Modern Persian, *-rā* appears with specific noun phrases in various positions. Some of these forms are still used in formal literary texts (FLT). They represent the benefactive *barā* “for” (42a) and (43a), *(be)suy* “towards” (44a), *az* “from, of” (45a), and *be* “to” (46a) in Modern Persian (MP). The modern version of each sentence immediately follows the CMP/FLT version.<sup>5</sup>

- (42) a. *amr rā hamē māl mē bāyist*  
 amar ra always stuff ASP-must-be.  
 “Amar always needs stuff” (Lazard 1963: 285)  
 Or: As for Amar, stuff is always needed.
- b. *māl hamishe barā-ye amr bāyad bash-ad*  
 stuff always For-EZ amar must be-3SG  
 “There must always be stuff for Amar.”
- (43) a. *pro in mehnat rā darmān-i andishide-am*  
 this suffering rā remedy-INDF thought-1SG  
 “As for this suffering, I have thought (of) a remedy.”
- b. *pro barā-ye in mehnat darmān-i andishide-am*  
 for-EZ this suffering remedy- INDF thought-1SG  
 [lit. “for this suffering I have thought of a remedy.”]
- (44) a. *man to rā ay-am*  
 I you rā come-1SG  
 “I will go to you” (Lazard 1963: 360)
- b. *man (be)suy-e to āy-am*  
 I (to) side-EZ you come-1SG  
 “I will come to/towards you.”
- (45) a. *loghmān rā porsid-and adab az ke āmuxt-i*  
 Loghman rā asked-3PL politeness from whom learned-2SG  
 “They asked (of) Loghman, whom did you learn politeness from.”
- b. *az loghmān porsid-and adab az ke āmuxt - i*  
 of Loghman asked-3PL politeness of whom learned-2SG  
 [lit. “(they) asked of Loghman from whom (you) learned politeness”]
- (46) a. *pro amir-rā zakhm-i zad-am*  
 king-rā wound-INDF hit-1SG  
 “As for the king, I wounded (him).”
- b. *pro be amir zakhm-i zad-am*  
 to king wound-INDF hit-1SG  
 [lit. “I inflicted a wound on the king.”]

5. We have made slight changes to the glossing of the borrowed data for the sake of consistency. Thanks to Mohsen Mahdavi for providing some of the FLT data in this section.

Note that the vocabulary choice in colloquial Modern Persian is different in some cases than that in the Classical Modern Persian or formal literary texts. However, for the sake of consistency, we are using the same vocabulary.

Relevant to our proposal is the fact that in all CMP/FLT cases, the DP+*rā* is c-commanded by an overt DP or *pro*. Thus it is assigned dependent case in Narrow Syntax, and is marked by *-rā* post-syntactically. Note that in (42b), the DP+*rā* originates in a position lower than the subject *māl*.

Possessive constructions manifest yet another version of the DP+*rā* in CMP. The MP equivalent of the CMP sentences clearly reveal the possessive construction hidden in the archaic forms.

- (47) a. *va pro in -rā nām shāhnāmeḥ nahād-and*  
 and this -rā name Shahname put-3.PL  
 “Its name they marked Shahname.”  
 [lit. “And this, they put the name Shahname on (it).”]
- b. *va pro [nām -e in] -rā Shāhnāmeḥ nahād-and*  
 and [name -EZ this] rā Shahnameh put-3.PL  
 “And its name they called Shahnameh.”
- (48) a. *pro xalgh-rā xun be-rixt-and*  
 people-rā blood SBJV-shed-3.PL  
 “As for people, they shed (their) blood.”
- b. *pro [xun - e xalgh] be-rixt-and*  
 blood-EZ people SBJV-shed-3.PL  
 [lit. “(they) shed people’s blood.”]

These cases, similar to the previous ones, are accounted for by the proposal at hand: The DP+*rā* is c-commanded by another overt or covert DP, assigned dependent case, and marked morphologically by *-rā* post-syntactically.

The morpheme *-rā* also appears in a different possessive construction represented by the example in (49a): *bud* “was” is a copula, yet *-rā* appears following the DP *u* “s/he” in the clause whose predicate is this copula. The modern version of this sentence is the one in (49b) where *-rā* is missing.

- (49) a. *u -rā pesar-i bud* (Lazard 1963: 191)  
 he-rā son-INDF was  
 “S/he had a son.”
- b. *u pesar-i dāsh*  
 he son-INDF had  
 “S/he had a son.”

It has been suggested that possessive constructions have an underlying HAVE, and that this element is in fact a preposition incorporated into the verbal BE (Harley 1995,

2002, among others). Benveniste (1966) noticed that many languages represent the possessive as a combination of *be* plus some spatial or locative preposition. Others, including Guéron (1995), Freeze (1992) and Kayne (1993) have proposed encoding this decomposition as part of UG, that is, to suggest that HAVE is represented as P in these constructions in all languages underlyingly. Thus those languages with verbal HAVE incorporate the P into the BE to produce the verb HAVE overtly.

Given this introduction, we propose the structure in (50) as the underlying structure for (49a), adopted from Harley (2002). The functional *v* with the flavor BE plus P representing HAVE provides a possessive interpretation. The DP *u* ‘s/he’ originates in the Specifier of the prepositional phrase. We suggest that this element moves out of PP, possibly into the Specifier of VP (similar to the situation in (23) and (25)), where it is c-commanded by an expletive *pro*, and therefore receives dependent case.<sup>6</sup>

(50) [ *pro* [<sub>VP</sub> *u*<sub>i</sub> [<sub>PP</sub> *t*<sub>i</sub> [<sub>P</sub> HAVE *pesar-i* ] ] *bud* ] ]

One question that arises is why these constructions do not exist in colloquial Modern Persian. As mentioned before, *-rā*, a residue of Old Persian *rādi* and Middle Persian *rāy* served several purposes in CMP, including Oblique/Dative/Benefactive cases, in addition to accusative case. However, this morpheme has lost its oblique function in Modern Persian. Relevant to our analysis is the point that those DPs were marked in the same syntactic construction as Modern Persian: That is, they were assigned dependent case via case competition with another, c-commanding DP. This analysis, if on the right track, suggests that in CMP, a DP c-commanded by another DP received dependent case as long as it was not embedded inside a prepositional phrase, similar to the situation in MP.

## 7. Concluding remarks

This article adopts a dependent case analysis of *-rā* in Persian, treating the morpheme as marking DPs that are both specific and assigned accusative case via case competition with a c-commanding DP in the syntax. The article also provides further support for treating nominative as unvalued for case, and makes use of this analysis to provide an explanation for the possibility of *rā*-marking on subjects raised from an embedded clause. Furthermore, this system accounts for all DP+*rā* cases, including direct objects, discussed in this chapter. It also explains why objects

---

6. Karimi (2005) suggests that Modern Persian lacks expletives. If this assumption is correct, the disappearance of expletives must be part of the syntactic change manifested in Modern Persian.



of prepositions are not marked by *-rā*, while DPs corresponding to the pronominal object clitic of P are.

We also argued that *-rā* marked a wider range of specific DPs in CMP, including dative/benefactive DPs in addition to accusative cases. The function of this morpheme, we claimed, is now restricted only to DPs valued exclusively for accusative case in MP.

The analysis adopted here implies that topic DPs, if merged directly in a topic position, are unvalued for case, and thus unmarked, similar to subjects. We saw that this claim is empirically supported.

There remains one case that provides a piece of counter evidence for the current analysis, at least at the first glance. The sentences in (51) allow *-rā* to mark the topic DP. In fact, the DP and the morpheme *-rā* are both obligatory in these cases. The complex predicate consists of the non-verbal element *xosh* ‘pleasant’ and the light verb *āy-ad* ‘comes’. The DPs *in rang / rang-hā* ‘this color, these colors’ are the subjects of the complex predicate *xosh āmadan* ‘to like’:

- (51) a. *mā-rā in rang xosh ā-yad*  
 us-*rā* this color pleasant come-3.SG  
 ‘This color is pleasant to us.’  
 [*lit.* ‘to us, this color comes pleasing.’]
- b. *mā-rā in rang-hā xosh ā-yand*  
 us-*rā* this color-PL pleasant come-3PL  
 ‘These colors are pleasant to us.’  
 [*lit.* ‘to us, these colors come pleasing.’]

Given the theoretical proposal advanced in this article, we can assume, rather justifiably, that the topicalized element is in fact in a derived position, as in (52). That is, it receives accusative case in its original position prior to moving into the topic position.

- (52) [<sub>TopP</sub> [*mā-rā*]<sub>i</sub> [<sub>TP</sub> *in rang* [<sub>vP</sub> *t<sub>i</sub> xosh ā-yad* ]]]

Thus the appearance of the DP+*rā* in (51a) and (51b) receives the same analysis here as in other cases: the initial DP receives dependent case in its base position, and is thus marked by *-rā* post-syntactically.

Yet there is another issue that might be problematic for our analysis. That is, the presence of *-rā* is obligatory in an elliptical construction, as demonstrated in (53).

- (53) [<sub>TopP</sub> *Ali-(ro)*]<sub>i</sub> *pro* [<sub>vP</sub> *t<sub>i</sub> fekr mi-kon-am* [<sub>CP</sub> *e<sub>i</sub> barande be-sh -e*, ]]]  
 Ali-*rā* thought ASP-do-1SG winner SBJV-become-3.SG  
 (*vali Maryam-\*(ro) pro* [<sub>vP</sub> *t ne- mi- dun -am*]  
 but Maryam-*rā* NEG-ASP-know-1SG  
 ‘As for Ali, I think he wins, but I don’t know about Maryam.’

The reason for the obligatory marking of the embedded subject might be due to certain properties of ellipsis. Rasekhi (2018) suggests that the contrasted element moves into the Specifier of Foc(us)P from its original position in order to escape the elided site. Thus the DP+*rā* in the second part of (53) must have received independent case in its original position. It can also be postulated that topic DPs can be base-generated in their surface positions, but not contrastively focused elements. We leave a thorough analysis of this matter, and a better understanding of ellipsis constructions, for future investigation.

## References

- Baker, Mark. 2017. Structural case: A realm of syntactic microparameters. In Simin Karimi & Massimo Piattelli-Montabelli (eds.), *Parameters, what are they, where are they? Special issue of Linguistic Analysis*. 41(3/4). 191–240.
- Baker, Mark & Nadya Vinokurova. 2010. Two modalities of Case assignment: Case in Sakha. *Natural Language & Linguistic Theory* 28. 593–642. <https://doi.org/10.1007/s11049-010-9105-1>
- Benveniste, Emile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.
- Bhatt, Rajesh. 2007. Unaccusativity and case licensing. Talk presented at McGill University.
- Browne, Wayles. 1970. More on definiteness marker: Interrogatives in Persian. *Linguistic Inquiry* 1. 359–563.
- Brunner, Christopher Joseph. 1977. *A syntax of western Middle Iranian* (Vol. 3). New York: Caravan Books.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In Michael J. Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.
- Diesing, Molly. 1992. Bare plural subjects and the derivation of logical representations. *Linguistic Inquiry* 23. 353–380.
- Enç, Murvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22. 1–25.
- Freeze, Ray. 1992. Existentials and other locatives. *Language*, 68:3. 553–595. <https://doi.org/10.2307/415794>
- Ghomeshi, Jila. 1996. Projection and inflection: A study of Persian phrase structure. Toronto: University of Toronto dissertation.
- Ghomeshi, Jila. 1997. Topics in Persian VPs. *Lingua*, 102(2). 133–167. [https://doi.org/10.1016/S0024-3841\(97\)00005-3](https://doi.org/10.1016/S0024-3841(97)00005-3)
- Guéron, Jacqueline. 1995. On have and be. In *Proceedings-NELS* 25. 191–206. Amherst: University of Massachusetts.
- Harley, Heidi. 1995. Subjects, events and licensing. Cambridge, MA: MIT dissertation.
- Harley, Heidi. 2002. Possession and the double object construction. *Linguistic Variation Yearbook* 2(1). 31–70. <https://doi.org/10.1075/livy.2.04har>
- Heim, Irene. 1982. The semantics of definite and indefinite noun phrases. Amherst, MA: University of Massachusetts dissertation.

- Holmberg, Anders & Urpo Nikanne. 2002. Expletives, subjects, and topics in Finnish. In Peter Svenonius (ed.), *Subjects, expletives, and the EPP*, 71–105. Oxford: Oxford University Press.
- Kahnemuyipour, Arsalan. 2014. Revisiting the Persian *Ezaf* construction: A roll-up movement analysis. *Lingua* 150. 1–24.
- Karimi, Simin. 1989. Aspects of Persian syntax: Specificity and the theory of grammar. Seattle, WA: University of Washington dissertation.
- Karimi, Simin. 1990. Obliqueness, specificity, and discourse functions. *Linguistic Analysis* 20(3/4). 139–191.
- Karimi, Simin. 1997. Persian complex verbs: Idiomatic or compositional. *Lexicology* 3(2). 273–318.
- Karimi, Simin. 2005. *A Minimalist approach to scrambling: Evidence from Persian*. Berlin: Walter de Gruyter. <https://doi.org/10.1515/9783110199796>
- Karimi, Simin & Mike Brame. 2013. A generalization concerning the *Ezaf* construction in Persian. *Linguistic Analysis* 38(1–2). 111–144.
- Kayne, Richard. 1993. Toward a modular theory of auxiliary selection. *Studia Linguistica* 47(1). 3–31. <https://doi.org/10.1111/j.1467-9582.1993.tb00837.x>
- Kornfilt, Jaklin & Omer Preminger. 2014. *Nominative as no case at all: An argument from raising-to-Accusative in Sakha*. Manuscript, Syracuse University.
- Lazard, Gilbert. 1963. *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.
- Mahootian, Shahrzad. 1997. *Persian*. London: Routledge.
- Marantz, Alec. 1991. Case and licensing. In German Westphal, Benjamin Ao & Hee-Rahk Chae (eds.), *Proceedings of the 8th Eastern States Conference On Linguistics (ESCOL 8)*, 234–253. Ithaca, NY: CLC Publications.
- Preminger, Omer. 2011. Agreement as a fallible operation. Cambridge, MA: MIT dissertation.
- Preminger, Omer. 2014. *Agreement and its failures*. *Linguistic Inquiry Monographs* 68. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/9780262027403.001.0001>
- Rāsekhi, Vahideh. 2018. Ellipsis and information structure: Evidence from Persian. Stony Brook, NY: Stony Brook University dissertation.
- Samiian, Vida. 1983. Structure of phrasal categories in Persian: An X-Bar analysis. Los Angeles, CA: University of California dissertation.
- Samiian, Vida. 1994. The *Ezaf* construction: Some implications for the theory of X-bar syntax. In Mehdi Marashi (ed.), *Persian studies in North America: Studies in honor of Mohammad Ali Jazayeri*, 17–42. Bethesda, MD: Iranbooks.

## The *Ezafe* construction revisited

Richard Larson and Vida Samiian

Stony Brook University / University of California, Los Angeles

This chapter addresses the nature, distribution and function of the *Ezafe* morpheme, a distinguishing grammatical feature of many of the Iranian languages. We review three main analyses advanced in the wide literature on the subject: semantic, morphological, and syntactic. We argue that the syntactic account of *Ezafe* is the most promising, both in its empirical reach, and explanatory power. Looking at an exhaustive range of data from Iranian Persian (iPersian) and other Iranian languages, we note that *Ezafe* occurs between nominal elements in the NP, PP, AP, and QPs. Following case theory (Chomsky 1981), we propose that *Ezafe* satisfies a licensing requirement in the following phrase, similar to ‘of’ in English. We then consider in detail the implications of this theory for the occurrence of *Ezafe* before PPs in iPersian and before finite and nonfinite complement clauses in iPersian and Kurdish. Finally, we examine the occurrence of *Ezafe* in Zazaki ‘double *Ezafe* constructions’ and in Caspian languages showing the so-called ‘Reverse *Ezafe* construction’ in light of the case-based analysis.

**Keywords:** Iranian languages, *Ezafe*, semantic approach, morphological approach, case-based analysis, licensing requirement, NP modifiers and complements, adjective phrases, Partitives, post-nominal PP modifiers, complement clauses, double *Ezafe* construction, Reverse *Ezafe*

### 1. Introduction

A distinguishing grammatical feature of many of the Iranian languages is the presence of the so-called ‘*Ezafe*’ morpheme (EZ), which is realized within a variety of phrases including nominals, prepositional phrases, adjectivals and quantifiers. Typical examples from Iranian Persian (from Samiian 1994) are given in (1), where *Ezafe* is realized as *ye/e* depending on whether the preceding form ends in a vowel or consonant (resp.):<sup>1</sup>

---

1. Abbreviations follow the list from Leipzig Glossing Rules.

- (1) a. [<sub>NP</sub> *xune=ye ma*] NP  
 House=EZ 1.PL  
 “our house”
- b. [<sub>NP</sub> *xune=ye kucik*] NP  
 house=EZ small  
 “small house”
- c. [<sub>PP</sub> *pošt=e divar*] PP  
 behind=EZ wall  
 “behind the wall”
- d. [<sub>AP</sub> *xeyli negæran=e bæce-ha*] AP  
 very worried=EZ child-PL  
 “very worried about the children”
- e. [<sub>QP</sub> *hæme/bištær=e xune-ha*] QP  
 all/most=EZ house-PL  
 “all/most of the houses”

An enduring question in Iranian grammatical studies is: what is the nature of the *Ezafe* morpheme? Where does it occur and why does it occur there? What conditions its distribution and what is its function? In the large literature on *Ezafe*, three general approaches have been pursued. One approach is broadly semantic: *Ezafe* serves to grammatically express or realize a semantic notion like modification or predication. A second approach is morphological: *Ezafe* is a morphological affix available only on a certain class of stems. The third approach is syntactic: *Ezafe* executes a specific function such as case-marking.

In this chapter we review the three main classes of proposals. We argue that although all face empirical challenges, the syntactic account of *Ezafe* appears the most promising, not only in terms of its empirical reach within Iranian ‘*Ezafe* languages’, but within languages showing the so-called ‘Reverse *Ezafe*’ construction as well, such as Gilaki, Mazanderani and Balochi. We begin with a brief review of the facts.

## 2. The *Ezafe* phenomenon

‘*Ezafe*’ refers to a morpheme occurring in Modern Persian,<sup>2</sup> Balochi, Kurdish (Sorani, Kurmanji), Zazaki (aka Dimili) and Gorani (including Hawrami). In these languages, N, A, Q and P heads precede their complements and modifiers.

---

2. Modern Persian comprises three main geographical variants spoken in Iran, Afghanistan and Tajikistan; while largely mutually intelligible, they are nonetheless linguistically distinct. In the text we refer to these variants as ‘iPersian’, ‘aPersian’ and ‘tPersian’, respectively, in deference to their speakers’ wishes to be identified as speaking Persian, but distinguishing them for linguistic purposes.

In certain cases, *Ezafe* (-EZ) appears between them, realized on the preceding element. The basic patterns are schematized in (2):

- (2) a. N - EZ NP/AP/PP/nonfinite CP  
 b. A - EZ  
 c. Q - EZ NP (for some Qs)  
 d. P - EZ NP (for some Ps)

iPersian exhibits *Ezafe* in its simplest form, the only variation being phonological (*el ye*). (3a)–(3g) show *Ezafe* occurring between a noun and a nominal complement or modifier. (3h) shows *Ezafe* between a noun and an attributive adjective. (3i) shows it between a noun and a PP. Finally, (3j) shows that *Ezafe* is recursive insofar as multiple attributive adjectives trigger multiple instances of it.

(3) **Modifiers & complements of Ns**

- a. *del=e sæng* (N=EZ NP)  
 heart=EZ stone  
 “stone heart”
- b. *mænzəl=e John* (N=EZ NP)  
 house=EZ John  
 “John’s house”
- c. *šæhr=e Tehran* (N=EZ NP)  
 city=EZ Tehran  
 “Tehran city”
- d. *Ali=e Ghozati* (N=EZ NP)  
 Ali=EZ Ghozati  
 “Ali Ghozati”
- e. *tæxrib=e šæhr* (N=EZ NP)  
 destruction=EZ city  
 “destruction of the city”
- f. *xordæn=e ab* (N=EZ NP)  
 drinking=EZ water  
 “drinking of water”
- g. *forunšænde=ye ketab* (N=EZ NP)  
 seller=EZ books  
 “seller of books”
- h. *otaq=e besyar kucik* (N=EZ AP)  
 room=EZ very small  
 “very small room”
- i. *divar=e jelo Ali* (N=EZ PP)  
 wall=EZ in-front-of Ali  
 “wall in front of Ali”

- j. *ketab=e sæbz=e jaleb* (N=EZ AP=EZ AP)  
 book=EZ green=EZ interesting  
 “interesting green book”

(4a)–(4c) illustrate the occurrence of *Ezafe* in an adjective phrase (AP) between the head and its nominal (NP) complement:

(4) **Complements of As**

- a. *ašeq=e Hæsæn* (A=EZ NP)  
 in love=EZ Hasan  
 “enamored with Hasan”
- b. *negæran=e bæce-ha* (A=EZ NP)  
 worried=EZ child-PL  
 “worried about the children”
- c. *montæzer=e Godot* (A=EZ NP)  
 waiting=EZ Godot  
 “waiting for Godot”

*Ezafe* also occurs in iPersian between some quantificational elements (Qs) and their restriction phrase (5):

(5) **Partitives**

- a. *tæmam =e šerkæt-ha* (Q=EZ NP)  
 all =EZ company-PL  
 “all/the-totality-of companies”
- b. *tæmam =e in šerkæt-ha* (Q=EZ NP)  
 all =EZ these company-PL  
 “all/the-totality-of these companies”
- c. *bištær =e in šerkæt-ha* (Q=EZ NP)  
 most =EZ these company-PL  
 “most/the-majority-of companies”

(6a)–(6d) illustrate an interesting alternation involving *Ezafe* and relative clauses (RCs). iPersian RCs are uniformly post nominal. Finite RCs (FRCs) do not involve *Ezafe* and are instead introduced by the relative marker *ke* (6a), (6b). By contrast, reduced, nonfinite RCs (RRCs) are introduced by *Ezafe* and no *ke* appears (6c), (6d):

(6) **Finite and reduced relative clauses**

- a. *dust =e Hæsæn] (\*=e) [ke Nanaz-o mi-šnas-e]*  
 friend =EZ Hasan =EZ that Nanaz-ACC DUR-know.PRES-3SG  
 “the friend of Hasan who knows Nanaz” (N FRC)

- b. *in šagerd-a (\*=ye) [ke zæbanšenasi mi-xun-ænd]*  
 DEM student-PL =EZ that linguistics DUR-study.PRES-3PL  
 “these students who study linguistics” (N FRC)
- c. *in jævan=e [æz swis bærgašt-e]* (N=EZ RRC)  
 this young=EZ from Switzerland re=turn.PST-PTCP  
 “this young man back from Switzerland”
- d. *æk=e [čap=šod-e dær ruzname]* (N=EZ RRC)  
 photo=EZ published-PTCP in newspaper  
 “the photo published in the newspaper”

Finally, (7a)–(7e) show that with certain iPersian PPs, *Ezafe* occurs between the P head and its object. (7f) shows furthermore that when such a PP occurs as a noun modifier, *Ezafe* may sometimes occur between PP and the head noun:

(7) Complements of (certain) Ps

- a. *beyn=e mæn=o to* (P=EZ NP)  
 between=EZ me=and you  
 “between you and me”
- b. *væsæt=e otaq* (P=EZ NP)  
 in-the-middle=EZ room  
 “in the middle of the room”
- c. *dor=e estæxr* (P=EZ NP)  
 around=EZ pool  
 “around the pool”
- d. *bæqæl=e dær* (P=EZ NP)  
 by=EZ door  
 “by the door”
- e. *kenar=e dærya* (P=EZ NP)  
 next=EZ sea  
 “on the beach”
- f. *xune=ye [kenar=e dærya]* (N=EZ [P=EZ NP])  
 house=EZ next=EZ sea  
 “house on the beach”



### 3. Analyses of *Ezafe*

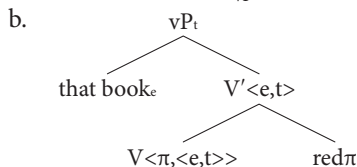
We now examine three approaches noted earlier – semantic, morphological and syntactic – considering their major assumptions and implications.

#### 3.1 Semantic analyses

##### 3.1.1 *Ezafe* as predication marker

Yadgar Karimi (2007) proposes that *Ezafe* instantiates a semantic predication relation between its two flanking expressions, developing ideas by Bowers (1993). Briefly, Bowers analyzes English predicates as property-denoting expressions of semantic type  $\pi$ , which require intervention by a predicate-forming operator (Pred) in order to be able to combine semantically with a subject. (8) illustrates the basic picture, where the semantic types of the various expressions are written as subscripts, and where we replace Bowers Pred(P) category with the more modern  $v(P)$  for ease of discussion in what follows.

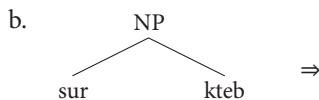
- (8) a. Mary considers [<sub>vP</sub> that book red ]

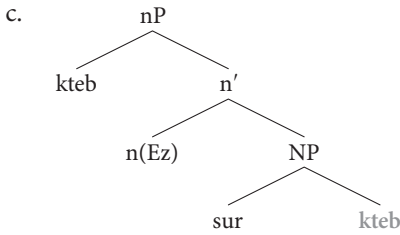


Here the AP ‘red’ is of type  $\pi$ , which is inappropriate for direct combination with ‘that book’, which is of type  $e$ . The predicate-forming operator  $v$  thus combines first, creating the  $v'$  predicate  $v'$ –‘red’, which now is of appropriate type  $\langle e,t \rangle$  to combine with the subject. The result is the (small) clausal expression ‘that book red’ of type  $t$ .

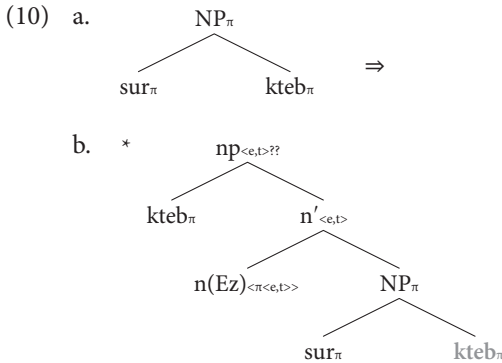
Karimi (2007) proposes to extend this picture to Kurdish *Ezafe* constructions, analyzing *Ezafe* as a predicate-forming operator counterpart to  $v$  above. Karimi offers the two-step derivation in (9b)–(9c) for the simple example in (9a), where *sur* ‘red’ initially merges with *kteb* ‘book’ as an adjunct, but where the latter ultimately raises to the Spec of a combining  $n/Ez$  head so that ‘predication should somehow be codified syntactically in the DP’ (Karimi 2007: 2164).

- (9) a. *kteb=i sur*  
 Book=EZ red  
 ‘(a) red book’





Karimi provides no type labels for the expressions in his trees, but it is plain that he sees a close analogy between structures (9c) and (8b). Thus he comments: “the merged  $n^0$  along with its F-selected complement constitutes a function with one unsaturated argument, the NP subject (the head N)” (2007: 2615). This assimilation is highly problematic, however. For Bowers (1993) *sur* and *kteb* would BOTH denote property expressions of type  $\pi$ , whose combination in (9b) would occur, not by predication, but rather by property conjunction. Furthermore, if one converted either *sur* or  $[_{NP} \textit{sur kteb}]$  to a propositional function ( $\langle e, t \rangle$ ), the result would be type-inappropriate for combining with *kteb*, which is not of type  $e$ , nor would the result be of type  $t$ , counterpart to (8):



The proposed analogy thus fails on closer inspection.

In our view, Karimi’s proposal rests on a basic semantic misunderstanding of examples like (9a), specifically of the semantic relation holding between its two constituents. Neither on a classic formal semantic analysis nor on one countenancing first-order properties as in Bowers (1993) is this relation predication. On a classic semantic analysis (e.g., Larson & Segal 1995; Heim & Kratzer 1998), *red* and *book* in (11a) are interpreted as predicates and their combination as co-predication – predication of the same object ( $x$ ) (11b). On Bowers’ property-theoretic analysis, they are conjoined properties (11c).

- (11) a.  $[_{NP} \textit{red book}]$   
 b.  $\lambda x[\textit{red}'(x) \ \& \ \textit{book}'(x)]$   
 c.  $\textit{red}' \cap \textit{book}'$

Hence despite what Karimi (2007) suggests, there is in fact no natural analogy between predication structures like (8b) and what is occurring in modification, and hence no natural analysis of *Ezafe* as a predicate-forming operator.

In addition to the formal points made above, we also take note of a set of examples due to Ghomeshi (p.c.), which are problematic for Karimi (2007) and other accounts appealing to predication.<sup>3</sup> It is well-known that whereas many adjectives are predicative (12a), some are not (13a). Correspondingly, combination with a noun is equivalent to co-predication (12c) or is not (13c):

- (12) a. John is elderly.  
 b. John is a friend.  
 c. John is an elderly friend. = John is a friend who is elderly.
- (13) a. #John is longstanding.  
 b. John is a friend  
 c. John is a longstanding friend. ≠ #John is a friend who is longstanding.

If *Ezafe* were associated strictly with predicative relations, as in Karimi (2007), one might expect it to occur strictly with predicative adjectives and not with non-predicative ones. This is not the case, however. As Ghomeshi (p.c.) notes, there is no difference between predicative and non-predicative adjective-noun combinations in Persian with regard to occurrence with *Ezafe*. If an appropriate adjective of either type exists, then *Ezafe* is present.

- (14) *Hæsæn dust =e mosen-i-e*  
 Hæsæn friend =EZ elderly-INDF-be.3SG  
 “Hasan is an elderly friend.” (co-predicational)
- (15) *Hæsæn dust =e qadimi-i-e*  
 Hæsæn friend =EZ longstanding-INDF-be.3SG  
 “Hasan is a longstanding friend.” (non-copredicational)
- (16) *Yoyo cellist =e xeyli xub-i-e*  
 Yoyo cellist =EZ very good-INDF-be.3SG  
 “Yoyo is a very good cellist.” (non-copredicational)

---

3. Examples like (15)–(19) also bear against analyses like Den Dikken & Singhapreecha (2004) and Den Dikken (2006) that attempt to construe *Ezafe* structures as instances of subject-predicate inversion, with *Ezafe* instantiating Pred. The relation between the nominal and the adjective in (15)–(19) is neither predication nor co-predication. Likewise these data also undermine approaches like Franco et al. (2015), which attempts to analyze *Ezafe* constructions semantically as a series of co-predications. Again, the relation between the nominal head and the predicate in (15)–(19) cannot be captured as co-predication.

- (17) *Clinton ræis-jomhur =e sabeq=e amrika-st*  
 Clinton president =EZ past=EZ America-be.3.sg  
 “Clinton is a past president of America.” (non-copredicational)
- (18) *Biden ræis-jomhur =e ayænde=ye amrika-st*  
 Biden president =EZ future=EZ America-be.3.sg  
 “Biden is a future president of America.” (non-copredicational)
- (19) *Alex modir =e kar.košte-i-e*  
 Alex manager =EZ veteran-INDF-be.3.sg  
 “Alex is a veteran manager.” (non-copredicational)

Hence assimilating *Ezafe* to predicative semantics seems both theoretically and empirically mistaken.

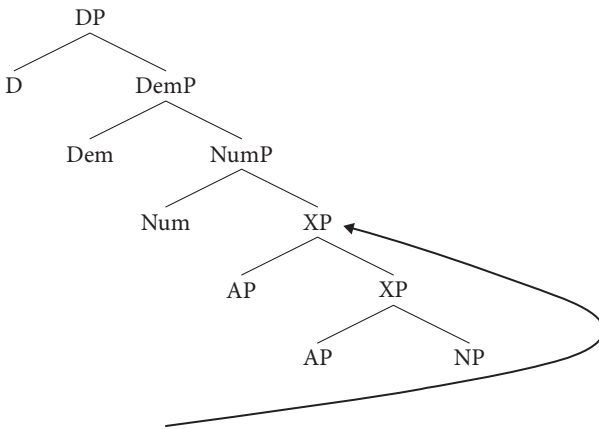
### 3.1.2 *Ezafe* as modification marker

Consider next the often-repeated description of *Ezafe* as a ‘marker of modification’.<sup>4</sup> A recent analysis embodying this claim is Kahnemuyipour (2014), who addresses iPersian. Although Kahnemuyipour offers no structural analysis of any actual iPersian example in his paper, two general diagrams that he provides allow one to reconstruct the basic idea. The analysis is presented in somewhat simplified form in (20a)–(20c). Briefly, Kahnemuyipour (2014) adopts the proposal of Cinque (2010) according to which there exists a universal, right-descending structure for DPs wherein determiners, demonstratives and numerals occur higher and modifiers like APs occur lower and closer to the NP head (20a). Kahnemuyipour’s proposal for *Ezafe* concerns the lower portion of this structure, approximately as in (20b) for the example *ketâb-e sabz-e jâleb* ‘interesting green book’. The APs *sabz* ‘green’ and *jâleb* ‘interesting’ are positioned within AgrPs as specifiers of their own phrases (XP, YP) whose category Kahnemuyipour does not identify. The head noun *ketâb* ‘book’ is positioned at the bottom. To derive the correct surface order, [<sub>NP</sub> *ketâb*] raises to the specifier of AgrP and Y raises to Agr, where the two heads are realized as *Ez*. In the next stage (20c) the entire lower AgrP raises to the Spec of the higher ArgP. X raises to the higher Agr, and the two are again realized as *Ez*. This ‘roll-up movement’ yields the surface order of terms. *Ezafe* is analyzed as realizing agreement.<sup>5</sup>

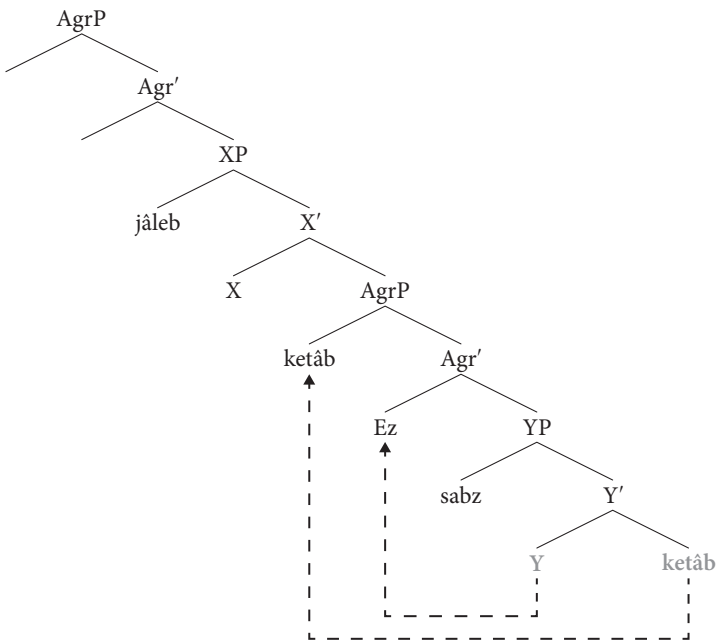
4. See for example, Palmer (1971) and Haig (2011).

5. Kahnemuyipour’s analysis of *Ezafe* as an agreement head is broadly plausible for AP modifier cases given that adjectives show agreement with NP in many languages. However the view loses plausibility with PP modifiers, which do not show agreement with Ns they modify

(20) a.



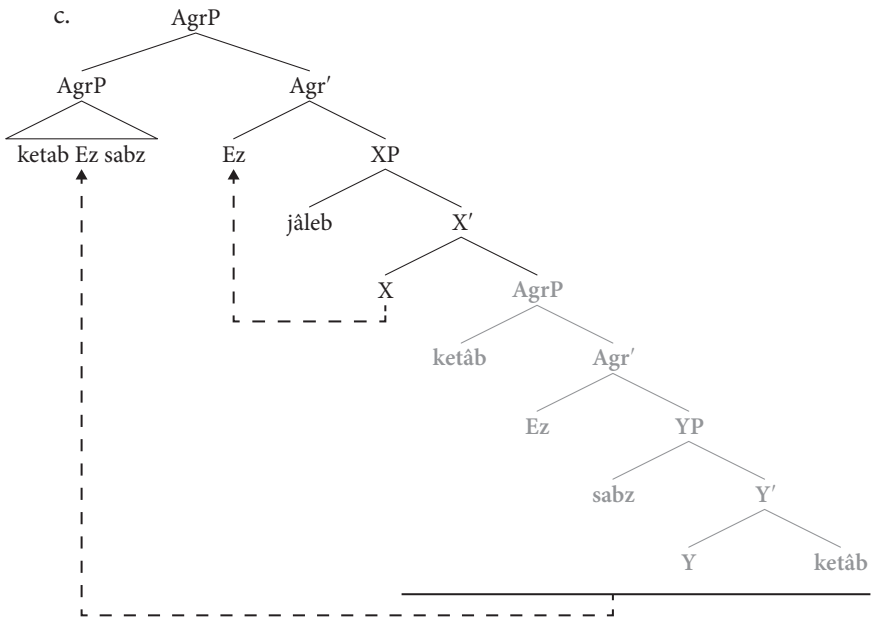
b.



cross-linguistically, but which nonetheless may show *Ezafê* in combination with an NP as shown in (i).

- (i) a. *Mina aks* <sup>\*(=e)</sup> [<sub>PP</sub> *dar ganje-râ*] *be* *Ali dâd*  
 Mina picture =EZ in closet-ACC to Ali gave  
 “Mina gave the picture in the closet to Ali.” (Kahnemuyipour 2014)
- b. [*sobh-hâ =ye* [<sub>PP</sub> *bâ kabutar-hâ*]-*râ*]  
 morning-PL =EZ with pigeon-PL-ACC  
 “the mornings with the pigeons.” (Samvelian, 2008)

Kahnemuyipour (2014) notes examples like (i) but provides no analysis of them within his account.



In Kahnemuyipour’s framework, examples with additional instances of *Ez* + AP modifier can be derived by additional *Agr* projections + roll up. Nonetheless, serious problems arise when a wider range of *Ezafe* cases is considered. Roll up movement as proposed by Cinque (2010) and Kahnemuyipour (2014) is uniformly PHRASAL movement. In (20b) and (20c), the items moving to *Agr* spec positions – *ketab* and *ketab Ez sabz* – are phrases in each case – NP and *AgrP*, respectively. It follows that Kahnemuyipour’s analysis of *Ezafe* will be unavailable when what needs to be fronted is a head. More specifically, it will be unavailable whenever the relation between  $\alpha$  and  $\beta$  in  $\alpha$  *Ez*  $\beta$  is a head-complement relation.

But we have already seen numerous instances of *Ezafe* occurring between heads and complements. Thus in (3e)–(3g) (repeated below), each of the nouns has a relational semantics ( $\lambda y \lambda x [R(y)(x)]$ ) and what follows *N* stands in a complement relation to it, supplying an argument to the relation, just as in the English gloss.

(3) *Ezafe* marking complements of *Ns*

e. *tæxrib=e šæhr* (N=EZ NP)

destruction=EZ city

“destruction of the city”

$\lambda y \lambda x \lambda e [\text{destruction}'(e)(y)(x)](\text{the-city}') \Rightarrow \lambda x \lambda e [\text{destruction}'(e)(\text{the-city}')](x)$

f. *xordan=e ab* (N=EZ NP)

drinking=EZ water

“drinking of water”

$\lambda y \lambda x \lambda e [\text{drinking}'(e)(y)(x)](\text{water}') \Rightarrow \lambda x \lambda e [\text{drinking}'(e)(\text{water}')](x)$

- g. *foruſænde=ye ketab* (N=EZ NP)  
 seller=EZ book  
 “seller of books”  
 $\lambda y \lambda x [\text{sell}(y)(x)](\text{books}') \Rightarrow \lambda x [\text{sell}'(\text{books}')(x)]$

Similarly in (4) (repeated below) each of the A's has a relational semantics and what follows A stands in a complement relation to it, as in the English gloss.

- (4) **Ezafe marking complements of As**
- a. *aſeq=e Hæſæn* (A=EZ NP)  
 enamored=EZ Hasan  
 “enamored with Hasan”/“in love with Hasan”  
 $\lambda y \lambda x [\text{enamored-of}'(y)(x)](\text{Hasan}') \Rightarrow \lambda x [\text{enamored-of}'(\text{Hasan}')(x)]$
- b. *negæran=e bæce-ha* (A=EZ NP)  
 worried=EZ child-PL  
 “worried about the children”  
 $\lambda y \lambda x [\text{worried-about}'(y)(x)](\text{child}') \Rightarrow \lambda x [\text{worried-about}'(\text{child}')(x)]$
- c. *montæzer=e Godot* (A=EZ NP)  
 waiting=EZ Godot  
 “waiting for Godot”  
 $\lambda y \lambda x [\text{waiting-for}'(y)(x)](\text{Godot}') \Rightarrow \lambda x [\text{waiting-for}'(\text{Godot}')(x)]$

In (7b)–(7e) (repeated below) each of the P's has a relational semantics and what follows P corresponds to its object, as in the English gloss.

- (7) **Ezafe marking complements of (certain) Ps**
- b. *væſæt=e otaq* (P=EZ NP)  
 in-the-middle-EZ room  
 “in the middle of the room”  
 $\lambda y \lambda x [\text{the-middle-of}'(y)(x)](\text{room}') \Rightarrow \lambda x [\text{the-middle-of}'(\text{room}')(x)]$
- c. *dor=e estæxr* (P=EZ NP)  
 around=EZ pool  
 “around the pool”  
 $\lambda y \lambda x [\text{around}'(y)(x)](\text{pool}') \Rightarrow \lambda x [\text{around}'(\text{pool}')(x)]$
- d. *bæqæl=e dær* (P=EZ NP)  
 by=EZ door  
 “by the door”  
 $\lambda y \lambda x [\text{by}'(y)(x)](\text{door}') \Rightarrow \lambda x [\text{by}'(\text{door}')(x)]$
- e. *kenar=e dærya* (P=EZ NP)  
 beside=EZ sea  
 “on the beach”  
 $\lambda y \lambda x [\text{on}'(y)(x)](\text{beach}') \Rightarrow \lambda x [\text{on}'(\text{beach}')(x)]$

Finally, under standard generalized quantifier semantics (Barwise & Cooper 1981), *šerkætha* “companies” in (5a)–(5b) (repeated below as (21a–b)) would be analyzed as supplying the restriction argument of the relational quantifier *tæmam* “all”. In no sense is *šerkætha* a modifier.

- (21) **Ezaf marking complements of partitive Qs**
- a. *tæmam=e šerkæt-ha* (Q=EZ NP)  
 all=EZ company-PL  
 “all/the-totality-of companies”  
 $\lambda Q\lambda P\forall x[Q(x) \rightarrow P(x)](\text{company}') \Rightarrow \lambda P\forall x[\text{company}'(x) \rightarrow P(x)]$
- b. *tæmam=e in šerkæt-ha* (Q=EZ NP)  
 all=EZ this company-PL  
 “all/the-totality-of these companies”  
 $\lambda Q\lambda P\forall x[Q(x) \rightarrow P(x)](\text{company}') \Rightarrow \lambda P\forall x[\text{company}'(x) \rightarrow P(x)]$

These cases pose a clear-cut challenge for Kahnemuyipour (2014). Since the relation here is uniformly head-complement, the observed ordering cannot be derived by phrasal roll-up; rather head movement would be needed (22)

- (22) a. **Head movement of N**
- $[_N \text{tæxrib}] = e \text{ } [_{NP} \text{šæhr } [_N \text{tæxrib}]]$   
 ↑ ..... ↓  
 =EZ city destruction  
 “destruction of the city”
- b. **Head Movement of A**
- $[_A \text{ašeq}] = e \text{ } [_{AP} \text{Hæsæn } [_A \text{ašeq}]]$   
 ↑ ..... ↓  
 =EZ Hasan enamored  
 “enamored with Hasan”
- c. **Head Movement of P**
- $[_P \text{bæqæl}] = e \text{ } [_{AP} \text{dær } [_P \text{bæqæl}]]$   
 ↑ ..... ↓  
 =EZ door by  
 “by the door”
- d. **Head Movement of Q**
- $[_Q \text{tæmam}] = e \text{ } [_{QP} \text{šerkætha } [_Q \text{tæmam}]]$   
 ↑ ..... ↓  
 =EZ companies all  
 “all companies”



But how this could work under Kahnemuyipour (2014) is unclear. Derivations strictly following (20b)–(20c) would require moving a head to a phrasal position, which is excluded under current theory. We see no plausible extension of Kahnemuyipour’s account to handle such cases, although one is clearly required.

Thus assimilation of *Ezafe* to structures of modification, as in Kahnemuyipour (2014), seems no more successful than assimilation of *Ezafe* to structures of predication, as in Karimi (2007). In both cases, the range of examples exhibiting *Ezafe* outstrips the single semantic concept that *Ezafe* is hypothesized to embody. Beyond trying to force iPersian *Ezafe* into a semantic space that is too narrow to accommodate it, the particular *Ezafe*-as-modifier analysis of Kahnemuyipour (2014), involving Cinque-style roll up, encounters technical problems that seem quite difficult to address.

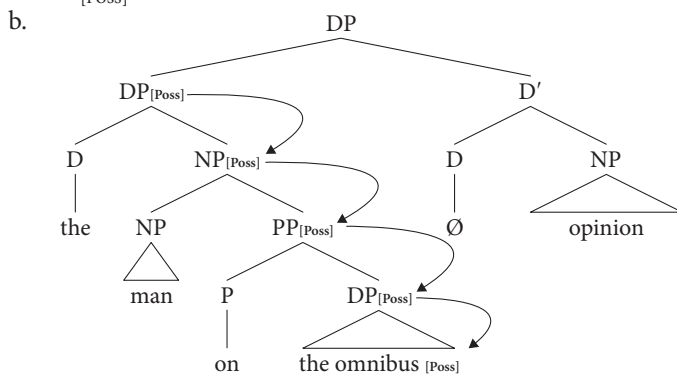
### 3.2 *Ezafe* as a morphological affix

Samvelian (2007, 2008) proposes a morphological analysis according to which *Ezafe* is an affix attaching to nominal elements (N, A or P) and marking a ‘dependency relation’ – modification, complementation, or possession – with its following phrase. Samvelian’s account is cast within Head-Driven Phrase Structure Grammar (HPSG) and crucially relies on the notion of EDGE features developed by Nevis (1986), Zwicky (1987), Lapointe (1990, 1992) and Miller (1991). The English prenominal (or ‘group’) genitive can be used to illustrate the approach, drawing on discussion from Anderson (2013). As Anderson notes, English prenominal genitive inflection is realized on a variety of items, from lexical words (23a) to much larger phrases (23b)–(23e). In all cases, however, it is realized at a right edge, whatever the categorial identity or grammatical function of the item it attaches to. ((23) = (1a)–(1e) in Anderson 2013):

- (23) a. *[Fred]’s opinion about the English genitive is different from mine.*  
 b. *[The man on the Clapham omnibus]’s opinion about the English genitive is poorly thought out.*  
 c. *[Every linguist I know]’s opinion about the English genitive involves functional categories.*  
 d. *[That young hotshot who was recently hired at Princeton that I was just telling you about]’s opinion about the English genitive is simply wrong.*  
 e. *Even [that colleague who shares an office with you]’s opinion about the English genitive is not to be trusted.*

In HPSG, this distribution can be captured by assigning a right-EDGE feature [POSS] to the possessor in a possessive DP. The nature of this feature is to propagate downward from mother node to its rightmost daughter until it reaches a terminal element where it is pronounced (24) (modeled on (6)–(7) in Anderson 2013):

- (24) a. **English Possessive**  
 Type: [EDGE: RIGHT]  
 Value: [POSS]  
 Word-level Morphology:  
 $/X_{[POSS]}/ \rightarrow /X+z/$



Samvelian (2007) extends this general picture to *Ezafe*, using a right edge feature [EZ] that can be affixed either to words or to NPs.

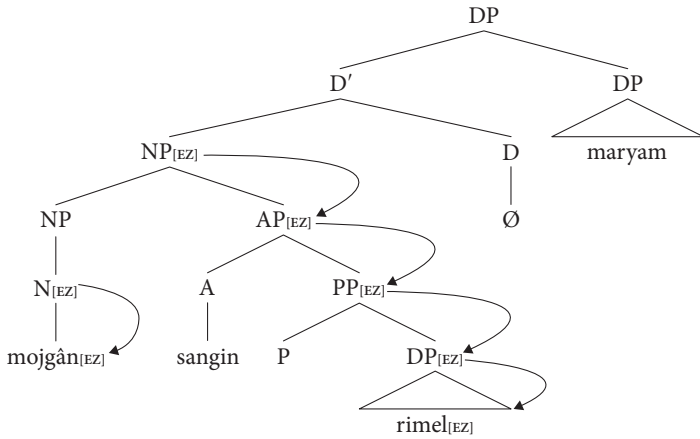
- (25) **iPersian Ezafe**  
 Type: [EDGE: RIGHT]  
 Value: [EZ]  
 Word-level Morphology:  $/X_{[EZ]}/ \rightarrow /X+e/$

To illustrate this proposal, consider (26) (adapted from (56) in Samvelian 2007), which differ in whether the PP *æz rimel* “of/with mascara” is positioned after or before the adjective *sængin* “heavy” (resp.). Note that *Ezafe* attaches to *rimel* or to *sængin* depending on word order:

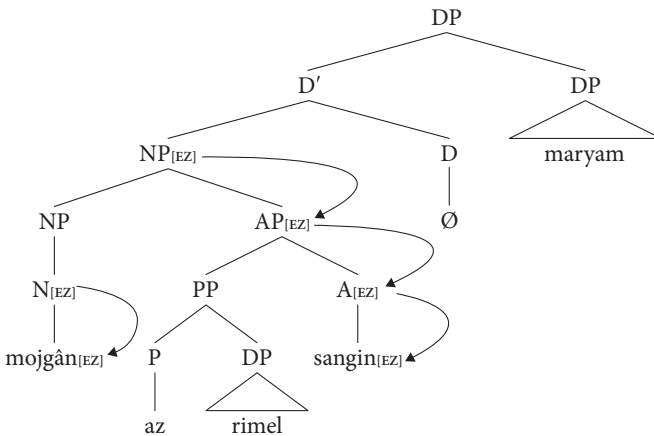
- (26) a. *mojgan=e sængin æz rimel=e Maryam*  
 eyelid.PL=EZ heavy of mascara=EZ Maryam  
 “eyelids heavy with mascara of Maryam’s”/“Maryam’s mascara-laden eyelids”
- b. *mojgan=e æz rimel sængin=e Maryam*  
 eyelid.PL=EZ of mascara heavy=EZ Maryam  
 “eyelids heavy with mascara of Maryam’s”/“Maryam’s mascara-laden eyelids”

Following the general idea in (24), this distribution can be captured by affixing *Ezafe* at the word level to the lexical *mojgan* “eyelid” and at the phrasal level to the NP *mojgan sængin æz rimel/mojgan æz rimel sængin* “eyelids heavy with mascara”, as shown in (27) (resp.). Note that in both trees [EZ] passes down a right edge starting from the high NP. Where [EZ] is realized depends on what is rightmost in the largest NP.

(27) a.



b.



Samvelian completes her 2007 account with an additional feature [DEP] (for ‘dependency’) that accompanies [EZ] and takes scope at the level where [EZ] attaches – i.e., at N or NP. As defined by Samvelian (2007: 636), [DEP] requires “that the constituent [it marks] must be followed by a noun, an AP, a PP or an NP.” Thus in (27) [DEP]-marking on the lexical noun [<sub>N</sub> *mojgân*] is satisfied by the presence of the AP [<sub>AP</sub> *sangin az rimel*]/[<sub>AP</sub> *az rimel sangin*]; [DEP]-marking on the noun phrase [<sub>NP</sub> *mojgân sangin az rimel*]/[<sub>NP</sub> *mojgân az rimel sangin*] is satisfied by the presence of “Maryam”.<sup>6</sup>

Samvelian’s account has advantages over the previous two insofar as it does not try to associate *Ezafe* with a single semantic concept like predication or

6. The structures in Samvelian (2007) do not include a DP projection in possessives as in (27). Furthermore, *Maryam* is classified as an NP rather than a DP. These modernizing adjustments in (27) do not appear to us to jeopardize the basic account.

modification. Instead, *Ezafe* serves as a general ‘sign of dependency’ – modification, complementation or possession – between the [EZ]-bearing nominal and the following phrase. At the same time, although Samvelian’s account yields a description of *Ezafe* distribution, it provides no explanation for it. *Ezafe* is proposed to be a nominal morpheme whose effect is to require a following AP, PP or NP: but if so:

- Why should nominal elements bear such marking?
- What unites the class of phrases selected by [DEP]? Do AP, NP and PP share some property such that they pattern together with respect to *Ezafe*; or are they simply a random list?

The force of these questions becomes clearer in the context of relative clauses and postnominal PP modifiers.

### 3.2.1 *Ezafe* and relative clauses

We observed earlier in (6) (repeated below) that whereas iPersian finite relative clauses (FRCs) resist *Ezafe* (6a)–(6b), iPersian reduced relative clauses (RRCs) require it (6c)–(6d).

- (6) a. *dust =e Hæsæn (\*=e) [ke Nanaz-o mi-šnas-e]* N FRC  
 friend =EZ Hasan =EZ that Nanaz-ACC DUR-know.PRES-3SG  
 “the friend of Hasan who knows Nanaz”
- b. *in šagerd-a (\*=ye) [ke zæbanšenasi mi-xun-ænd]* N FRC  
 DEM student-PL =EZ that linguistics DUR-study.PRES-3PL  
 “these students who study linguistics”
- c. *in jævan=e [æz Suis bær-gæšt-e]* (N=EZ RRC)  
 this young=EZ from Switzerland re-turn.PST-PTCP  
 “this young man back from Switzerland”
- d. *æks=e [čap=šod-e dær ruzname]* (N=EZ RRC)  
 photo=EZ publication=got-PTCP in newspaper  
 “the photo published in the newspaper”

Reduced relatives are not listed by Samvelian (2007) in the set of categories selected by her [DEP] feature” the categories that can follow *Ezafe*. Samvelian could, of course, modify the definition of [DEP] to include RRCs. But what explains their divergent behavior from FRCs? Is this an idiosyncratic fact or is it principled?

Relatedly, Samvelian (2008) notes that the Kurdish language Kurmanji differs from iPersian in permitting *Ezafe* before FRCs (28).<sup>7</sup>

7. Samvelian (2008) cites the Kurdish language Sorani as also allowing *Ezafe* before FRCs. We discuss Sorani and other relative clause data in §. 3.4.1

- (28) a. *mirov=ê* [ku min dît-î ]  
 Man=EZ.MASC.SG that I.OBL see-PAST  
 “the man that I saw” (Kurmanji, Samvelian 2008: 347)
- b. *çîrok=a* [ku wî ji min re got ]  
 story=EZ.FEM that 3S.OBL ADP 1S.OBL ADP say.PST.3S  
 “The story that he told me” (Kurmanji, Songül Gündoğdu p.c.)

Again, Samvelian could modify the definition of [DEP] for Kurmanji to include FRCs along with NPs, APs and PPs. But what explains the different behavior of iPersian vs. Kurmanji? Is this idiosyncratic variation to be listed, or does it trace to something systematic?<sup>8</sup>

A potential answer, to which we will return in detail in the next section, is suggested by additional relative clause examples from iPersian (29a)–(29b) and from Kurmanji (29c).<sup>9,10</sup>

---

8. We note that Kahnemuyipour’s (2014) analysis also yields no clear account of variation in RCs either within iPersian or across Iranian. On the roll-up account *Ezafe* would apparently be generated by head movement of X to Agr with RRCs but not with FRCs. What predicts this? And what is different with respect to (ib) in Kurmanji?

- (i) a. NP Ez RRC X NP  
 ↑-----↓  
 b. NP \*Ez FRC X NP  
 ↑-----↓

9. Some iPersian speakers prefer variants of (29b) with two FRCs, e.g. (i). However (29b) is also acceptable.

- (i) *dust=e javan-i [ke molaqat=kard-i] [ke as Swis bargashte]*  
 Friend=EZ youth-def that meeting=did-2SG that from Swiss returned  
 “a friend of the young man that you met that recently returned from Switzerland”

10. Songül Gündoğdu reports that Kurmanji speakers accept (30); but they regard (i), where *Ezafe* attaches to an overt pronominal element, as more natural:

- (i) *çîrok=a [ku wî ji min re got] ew-a [ku di rojnamê da derket]*  
 Story=EZ.F that 3S.OBL ADP 1S.OBL ADP say.PST.3S. EZ.F that ADP  
 newspaper.OBL PART come out.PST.3SG  
 “The story that he told me that was published in the newspaper”

The status of the pronominal element in *ew-a* is unclear to us. Gündoğdu (p.c.) suggests it might be an instance of so-called ‘demonstrative/anaphoric *Ezafe*’ (Haig 2011). If so the gloss of (i) is actually closer to “The story that he told me, the one that was published in the newspaper”. If ‘one’ takes ‘story-published-in-the-newspaper’ as its antecedent, this will be equivalent to the standard interpretation of recursive RCs as expressing successive intersection.

- (29) a. *jævan=e* [æz Swis bær-gæšt-e]=ye  
 young.man=EZ from Switzerland re-turn.PST-PTCP]=EZ  
 [estexdam=šod-e dær vezaræt=e færhæng] (iPersian)  
 employment=got-PTCP in ministry=EZ education  
 “the young man back from Switzerland employed by the Ministry of  
 Education”
- b. *dust=e* [æz Swis bærgæšte]=ye [jævan-i [ke molaqat=kærd-i]]  
 friend=EZ from Swiss returned=EZ youth-INDF that meet.PST-2SG  
 “the recently returned friend from Switzerland of the young man that you  
 met” (iPersian)
- c. *çïrok=a* [ku wî ji min re got ] ya [ku di  
 story=EZ.F that 3S.OBL ADP 1S OBL ADP say.PST.3S. EZ.F that ADP  
 rojnamê da derket ]  
 newspaper.OBL PART came3SG  
 “The story that he told me that was published in the newspaper”  
 (Kurmanji, Songül Gündoğdu p.c.)

In (29a)–(29b) we see that *Ezafe* not only precedes RRCs in iPersian, it follows them as well, here appearing after the participle *bargašte* “returned”. In (29c) we see that *Ezafe* not only precedes FRCs in Kurmanji, it also follows them, appearing after the finite verb *got* “say.PST.3s”. Samvelian’s phrasal affix analysis crucially assumes that *Ezafe* is NOMINAL morphology; i.e., whether [EZ] combines with a lexical word or with NP, passing down its right edge, [EZ] must be realized on a NOMINAL stem. (29a)–(29c) therefore imply for Samvelian that iPersian participles and Kurmanji finite verb complexes occurring inside relative clauses must be fundamentally nominal in character.<sup>11</sup> This suggests the following key generalization argued for explicitly by Samiiian (1983, 1994) and Karimi & Brame (1986, 2012),<sup>12</sup> but which Samvelian’s analysis neither states nor captures:

Generalization 1: *Ezafe* occurs between nominal elements.

If *Ezafe* occurs on nominal stems and whatever *Ezafe* can precede it can also follow, then this amounts to saying *Ezafe* always occurs BETWEEN nominals.

11. Anderson (2013) notes that what separates a phrasal affix analysis of a morpheme X from an analysis of X as a clitic is precisely whether X exhibits selectivity in the stems it attaches to. Anderson argues that an analysis of the English genitive morpheme ’s as in (24) above is mistaken precisely because ’s exerts no constraints on the stems it affixes to. Under this reasoning, Samvelian’s analysis of *Ezafe* as a phrasal affix is justified to the extent that *Ezafe* is selective in the relevant sense: that it attaches to nominal stems.

12. Samiiian (1983, 1994) and Karimi & Brame (1986, 2012) develop this idea in different ways. The former pursues the “neutralization hypothesis” of van Riemsdijk.

### 3.2.2 *Ezafe* and postnominal PP modifiers

Consider next *Ezafe* distribution with respect to post-nominal PP modifiers. As noted by Samiiian (1983, 1994), iPersian prepositions appear to divide into three distinct classes in relation to their complements; there are Ps that FORBID *Ezafe* before their object (Class 1), Ps that ALLOW *Ezafe* before their object (Class 2), and Ps that REQUIRE *Ezafe* before their object (Class 3).

#### (30) Class 1 Ps (forbid *Ezafe*)

- a. *æz* (\*=*e*) *Maryam*  
from (=EZ) *Maryam*  
“from *Maryam*”
- b. *ba* (\*=*ye*) *Hæsæn*  
with (=EZ) *Hasan*  
“with *Hasan*”
- c. *be* (\*=*ye*) *Ali*  
to (=EZ) *Ali*  
“to *Ali*”
- d. *dær* (\*=*e*) *Maryam*  
in/at/on (=EZ) *Maryam*  
“in/at/on *Maryam*”

#### (31) Class 2 Ps (allow *Ezafe*)

- a. *bala* (=ye) *divar*  
up (=EZ) wall  
“up the wall”
- b. *jelo* (=ye) *Hæsæn*  
in front (=EZ) *Hasan*  
“in front of *Hasan*”
- c. *ru* (=ye) *miz*  
on (=EZ) table  
“on top of the table”
- d. *tu* (=e) *divar*  
inside (=EZ) wall  
“inside the wall”

#### (32) Class 3 Ps (require *Ezafe*)

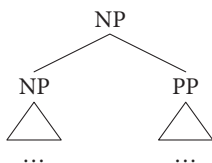
- a. *beyn* \*(=*e*) *mæn-o to*  
between =EZ me-and you  
“between you and me”
- b. *væsæt* \*(=*e*) *otaq*  
in-the-middle =EZ room  
“in the middle of the room”

- c. *dor* \*(=e) *estæxr*  
 around =EZ pool  
 “around the pool”
- d. *bæqæl* \*(=e) *dær*  
 by =EZ door  
 “by the door”

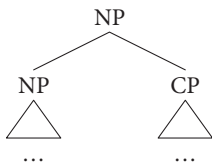
Samiian (1994) labels Class 1 Ps “True Prepositions” and she labels Class 2/Class 3 Ps “Nominal Prepositions”. *Ezafē* is licensed with the latter set because of the nominal status of the P to which -Ez attaches.<sup>13</sup>

iPersian PPs can, like iPersian relative clauses, function as adjunct modifiers of nominals, plausibly with a structure in (33a) (compare (33b)).

(33) a. **Modifying PP**



b. **Modifying RC**



Examples of adjunct PP modifiers headed by Ps of different classes are given in (35).

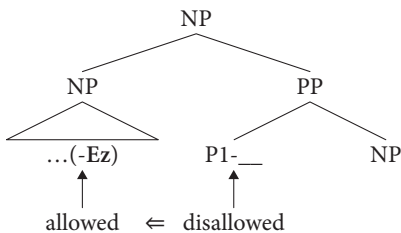
- (34) a. *šam* (=e) *ba* *Hæsæn* Class 1 P: *ba*  
 dinner =EZ with Hasan  
 “dinner with Hasan”
- b. *divar* =e *jelo* *Ali* Class 2 P: *jelo*  
 wall =EZ in-front-of Ali  
 “wall in front of Ali”
- c. *miz* =e *bæqæl* =e *Hæsæn* Class 3 P: *bæqæl*  
 table =EZ near =EZ Hasan  
 “table near Hasan”

13. Samiian (1983, 1994) says little about the source of alternation in Class 2 Ps; we return to this important issue in the next section.

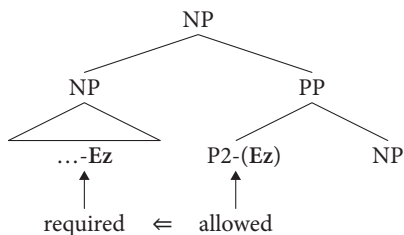


Note that in (34) *Ezafe* occurs not only inside PP, as determined by the class of its P head, in some cases it also occurs OUTSIDE PP on the modified nominal. Interestingly, presence of *Ezafe* on the nominal correlates with the class of preposition heading PP. Specifically, with Class 1 Ps, *Ez* is ALLOWED on the preceding NP (35a). With Class 2 Ps, *Ez* is REQUIRED on the preceding NP (35b). And with Class 3 Ps, *Ez* is REQUIRED on the preceding NP (35c).

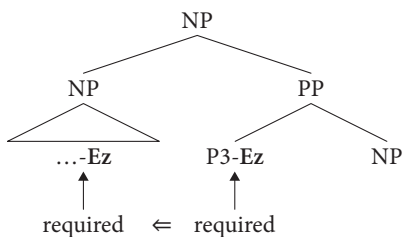
## (35) a. -EZ and P1's



## b. -EZ and P2's



## c. -EZ and P3's



These correlations are illustrated by the examples in (36):

- (36) a.  $[[_{NP} \textit{\text{šam}}] (=e) [_{PP} \textit{ba Hæsæn}]]$                       Class 1 P: *ba*  
           dinner (=EZ)                      with Hasan  
           “dinner with Hasan”
- b.  $[[_{NP} \textit{divar}] *(=e) [_{PP} \textit{jelo Ali}]]$                       Class 2 P: *jelo*  
           wall (=EZ)                      in-front-of Ali  
           “wall in front of Ali”

- c.  $[[_{NP} \textit{divar}] * (=e) [_{PP} \textit{jelo} =ye \textit{Ali}]]$  Class 2 P: *jelo*  
 wall =EZ in-front-of =EZ Ali  
 “wall in front of Ali”
- d.  $[[_{NP} \textit{miz}] * (=e) [_{PP} \textit{bæqæl} =e \textit{Hæsæn}]]$  Class 3 P: *bæqæl*  
 table =EZ near =EZ Hasan  
 “table near Hasan”

(36a) shows Class 1 *ba*, which forbids a following *Ezafe*; here *Ezafe* on the preceding nominal is optional. (36b) shows Class 2 *jelo* with no PP-internal *Ezafe*; here *Ezafe* is required on the preceding NP. (36c) shows *jelo* again, but with *Ezafe* present within PP; again *Ezafe* is obligatory on  $[_{NP} \textit{divar}]$  “wall”. Finally (36d) shows Class 3 *bæqæl* with obligatory internal *Ezafe*; PP-external *Ezafe* is required on the preceding NP.

The pattern in (35) – specifically (35b)–(35c) – presents a serious problem for Samvelian (2007). Her account allows for a nominal  $\alpha$  bearing *Ezafe* to require a FOLLOWING NP, AP, PP, etc. (37a). But it has no mechanism whereby a PP with certain properties can require a PRECEDING *Ezafe* (37b).

- (37) a.  $\alpha\text{-[EZ]} \longrightarrow \{\text{NP, AP, PP, RRC}\}$   
*requires*
- b.  $\alpha\text{-[EZ]} \longleftarrow [_{PP} \text{P=[EZ] NP}]$   
*requires*

In other words, given that (38a) is a possible iPersian NP-PP modifier structure, with neither NP nor P bearing *Ezafe*, nothing in Samvelian’s account as it stands will rule out (38b), where again NP and P bear no *Ezafe*. As (36b) shows, this structure is ungrammatical; *divar* requires *Ezafe*.

- (38) a.
- ```

      NP
     /  \
    NP   PP
   /  \  /  \
  sham ba Hæsæn
  
```
- b.
- ```

      *
      NP
     /  \
    NP   PP
   /  \  /  \
  divar jelo Ali
  
```

We believe that (37b) demonstrates a second fundamental generalization about *Ezafe* that any adequate account must capture, viz., that *Ezafe* is not present simply to signal the occurrence of a following phrase of a certain sort, as in Samvelian (2007); it is there to satisfy a “need” in that phrase.

Generalization 2: *Ezafe* satisfies a licensing requirement in the following phrase.

What (35b)–(35c) and (37b) suggest is that PPs headed by certain Ps have some requirement that a preceding *Ezafe* can discharge and without which the structure is ill-formed. This points once again, in our view, to the need to understand the class of phrases co-occurring with *Ezafe* – what unites them and what *Ezafe* supplies for them. We now turn to a theory that appears to have the right properties.

### 3.3 *Ezafe* as a case-marker

Consider the sets of examples below, involving NPs (39), APs (40), PPs (41) and QPs (42). In each, the (a) examples exhibit *Ezafe*; the remaining ones show either the iPersian preposition *æz* or *Ezafe* and *æz* alternating, with virtually identical meaning. Semantic variation across the example sets suggests that *æz* contributes very little on its own – i.e., that its content is determined contextually.<sup>14</sup> Like *Ezafe*, *æz* seems to be present largely for grammatical reasons, with examples becoming sharply ungrammatical without it.

- (39) a. *ye goruh =e/æz danešju-yan* NP  
 a group =EZ/of student-PL  
 “a group of the students”  
 b. *ye bæste =ye/æz ketab-ha-ye zæbanšenasi resid*  
 a package =EZ/of book-PL-EZ linguistics arrive.PST  
 “a package of books about linguistics arrived.”  
 c. *gozareš =e/æz vezaræt-e færhæng*  
 report =EZ/of ministry-EZ education  
 “report of/from the ministry of education”
- (40) a. *negæran =e bæce.ha* AP  
 worried =EZ child.PL  
 “worried about the kids”  
 b. *deltæng æz zendegi*  
 depressed of life  
 “depressed about life”  
 c. *xašmgin æz nætije =ye entexabat*  
 enraged of result =EZ election  
 “enraged by/at/about the election result”

14. iPersian *az* does have a contentful use as an ablative preposition meaning “from”. This use is also found with English ‘of’ in examples like (i). iPersian speakers detect ablative meaning with *az* in some of (39)–(42), for example (39c).

- (i) a. Alice jumped out of/from the plane.  
 b. Max ran out of/from the house.

- (41) a. *dær-tul =e mah =e Fevriye* PP  
 during =EZ month =EZ February  
 “during the month of February”  
 b. *qæbl =e/æz nahar*  
 before =EZ/of lunch  
 “before lunch”  
 c. *bæd =e/æz molaqat =e Hasan*  
 after =EZ/of visit =EZ Hasan  
 “after the meeting with Hasan”
- (42) a. *bištær =e ketab-ha* QP  
 most =EZ book-PL  
 “most of/among the books”  
 b. *bæzi æz ketab-ha*  
 some of book-PL  
 “some of/among the books”  
 c. *cænd=ta =ye/æz anha*  
 few=unit =EZ/of them  
 “few of them”  
 d. *hic kodum =ye/æz anha*  
 not any =EZ/of them  
 “none of them”

English exhibits the same broad parallelism between *Ezafe* and *æz* insofar as English can often gloss iPersian *Ezafe* quite naturally with ‘of’, its *æz*-equivalent in these contexts (43).<sup>15,16</sup> Here again the semantic contribution by ‘of’ is minimal. The preposition seems to be present for purely grammatical reasons.

15. The close parallelism between iPersian *Ezafe* and English *of* is noted explicitly in Karimi & Brame (1986, 2012) and Samiiian (1983, 1994). Note that Kahnemuyipour’s (2014) roll-up analysis would appear to require an entirely different treatment of *az* and *Ezafe*, since the latter is supposedly a manifestation of agreement whereas the former is a preposition. Likewise Samvelian’s (2007) analysis would appear to make the parallels accidental.

16. The status of *Ezafe* seems particularly clear in Northern Kurdish (Kurmanji) where nominals following *Ezafe* are overtly inflected for oblique case, exactly as they would be following a preposition; cf. (i):

- (i) a. =e *min*  
 ez 1SG.OBL  
 “mine” (e.g., *çav-e min* “my eye”/“eye of mine”, Pikkert 2010)  
 b. *ji min*  
 from 1SG.OBL  
 “from me”

Curiously, Franco et al. (2015) interpret the presence of oblique case-marking on Kurmanji nominals following *Ezafe* as evidence AGAINST the case-marking hypothesis. Presumably, however, the

- (43) a. *del=e sæng* N=EZ NP  
Heart=EZ stone  
“heart of stone”/“stone heart”
- b. *mænzel=e John* N=EZ NP  
house=EZ John  
“house of John’s”/“John’s house”
- c. *šæhr=e Tehran* N=EZ NP  
city=EZ Tehran  
“city of Tehran”/“Tehran city”
- d. *Ali=e Ghozati* N=EZ NP  
Ali=EZ Ghozati  
“Ali of the Ghozati’s”/“Ali Ghozati”
- e. *tæxrib=e šæhr* N=EZ NP  
destruction=EZ city  
“destruction of the city”
- f. *xordæn=e ab* N=EZ NP  
drinking=EZ water  
“drinking of water”
- g. *forunšænde=ye ketab* N=EZ NP  
seller=EZ book  
“seller of books”
- h. *bištær=e ketab-ha* Q=EZ NP  
most=EZ book-PL  
“most of the books”
- i. *arezumænd=e šohræt* A=EZ NP  
desirous=EZ fame  
“desirous of fame”
- j. *birun=e pænɟære* P=EZ NP  
out=EZ window  
“out of the window”
- k. *ba=vojud=e Hasan* P=EZ NP  
with=existence=EZ Hasan  
“in spite of Hasan”
- l. *be=dalil=e in mozu* P=EZ NP  
for=reason=EZ this issue  
“because of this issue”

---

same reasoning should apply to (ib): case-marking on nominals following prepositions should constitute evidence AGAINST the idea that the latter case-mark their objects. This is virtually a *reductio ad absurdum* in our view.

Chomsky (1981) proposes that “of” is present in the English expressions given as glosses in (43) in order to satisfy a case licensing requirement on NPs (i.e., on [+N] elements). In essence, nominal items require case, but nominal elements do not assign or check case. It follows that when two nominals X, Y are adjacent (44a), a case assigner like “of” will be required between them (44b) to assign case to the rightward Y. iPersian *æz* “of” can be analyzed in the same terms as (44c):

- | (44) | NON-CASE-ASSIGNING | CASE-ASSIGNING | CASE-REQUIRING                                   |
|------|--------------------|----------------|--|
| a.   | X[+N]              | ⇒              | Y[+N]  |
| b.   | X[+N]              | ⇒              | [ <sub>pp</sub> of ⇒ Y[+N] ] <b>English “of”</b> |
| c.   | X[+N]              | ⇒              | [ <sub>pp</sub> æz ⇒ Y[+N] ] <b>iPersian æz</b>  |

Samiian (1994) proposes essentially the same picture for iPersian *Ezafe*, suggesting that -EZ is a case-assigning element that is merged into the first nominal X and provides case assignment for the second nominal Y (45a). Larson & Yamakido (2008) offer a minor variant of this picture wherein *Ezafe* is, in effect, a clitic version of *az*, heading its own phrase (EzP) and cliticizing onto the preceding nominal stem (45b):

- | (45) | NON-CASE-ASSIGNING | CASE-ASSIGNING | CASE-REQUIRING  |
|------|--------------------|----------------|---|
| a.   | X[+N] - EZ         | ⇒              | Y[+N] <b>iPersian Ezafe</b>                           |
| b.   | X[+N] - EZ         |                | [ <sub>EzP</sub> - EZ ⇒ Y[+N] ] <b>iPersian Ezafe</b> |
|      | ↑                  | -----          | ↓   |

Samiian’s case-marking proposal (on either variant) directly accounts for the two key generalizations noted earlier, viz.:

Generalization 1: *Ezafe* occurs between nominal elements.

Generalization 2: *Ezafe* satisfies a licensing requirement in the following phrase.

Both generalizations derive from case theory as discussed – from the inability of nominal items to assign (or check) case and the licensing requirement on nominal elements that they receive case (or have it checked on them) and from the problem posed by adjacent nominals (44a).

### 3.4 Predictions of the case-marking analysis

The case-marking analysis makes a range of interesting predictions that distinguish it sharply from the three accounts reviewed above.

### 3.4.1 *Relative and complement clauses*

We noted earlier that iPersian *Ezafe* is unavailable before finite relative clauses (FRCs) (46a) but is required before reduced ones (RRCs) (46b).

- (46) a. *in jævan* (\*=*e*) [*ke æz Swis bærgæšte-æst*] (N FRC)  
 This youth =EZ that from Swiss returned-be.3sg  
 “this youth who has returned from Switzerland”
- b. *in jævan* =*e* [*æz Swis bærgæšt-e*] N-EZ RRC  
 this youth =EZ from Swiss return-PTCP  
 “this youth returned from Switzerland”

Neither Karimi (2007) nor Kahnemuyipour (2014) account for this divergence. Since FRCs and RRCs are both predicates semantically, no differences are expected under the first account. And since FRCs and RRCs are both modifiers semantically, no differences are expected under the second account. Under either, why should finite predicates/modifiers behave differently than reduced ones? Furthermore, we saw that Samvelian (2007, 2008) simply lists the categories that can follow *Ezafe* in her [DEP] feature, which offers no explanation for the difference in (46a)–(46b). Why should non-finite RCs be marked as “dependents” and not finite ones, given that the dependency relation is the same in both cases: attributive modification?

Similar results hold with iPersian complement clauses. *Ezafe* is unavailable in nominals before finite clausal complements (FCCs) (47a), but required before their reduced counterparts (RCCs) (47b).

- (47) a. *in omid* (\*=*e*) [*ke Shah æz Iran xahæd=ræft*] (N FCC)  
 this hope =EZ that Shah from Iran will-go  
 “the hope that the Shah will leave Iran.”
- b. *in omid* \*(=*e*) [*ræftæn-e Shah æz Iran*] N-EZ RCC  
 this hope =EZ go.INF-EZ Shah from Iran  
 “the hope of/for the Shah’s leaving Iran.”

Karimi (2007) and Kahnemuyipour (2014) make no predictions about this difference since complement clauses are neither predicates nor modifiers semantically, but rather arguments. In particular, their analyses cannot relate the absence/presence of *Ezafe* in (47) to its absence/presence in (46), despite the evident shared feature of finiteness/non-finiteness, respectively. And again, although Samvelian (2007, 2008) could certainly omit finite CPs in her [DEP] feature specification for *Ezafe* while including non-finite XPs, this provides no explanation for the difference. Why is “dependency” expressed by *Ezafe* with the one kind of propositional complement but not with the other? And why do relative and complement clauses pattern similarly across the two different dependency-types: modification vs. complementation?

By contrast, the case-marking analysis yields clear predictions in this domain. Assuming *Ezafe* is nominal morphology (or a nominal clitic) present to satisfy a case-marking requirement on the following phrase, its appearance is expected accordingly. Specifically:<sup>17</sup>

Prediction 1: *Ezafe* should occur BEFORE a clausal projection XP, if XP has nominal status (48a).

Prediction 2: *Ezafe* should ATTACH TO a clausal projection XP, if XP's final element  $\alpha$  is a nominal (48b).

(48) **Ezafe with clausal projections XP**

- |                        |  |
|------------------------|--|
| a. [ $\alpha$ -EZ XP ] | b. [ [ <sub>XP</sub> ... $\alpha$ ] -EZ YP ] |
| [+N]                   | [+N]    [+N]                                 |

We examine these predictions in detail below.

### 3.4.1.1 Prediction 1: *iPersian*

Traditional grammar refers to clausal complements as ‘noun clauses’ and to relative clauses as ‘adjectival clauses’ in virtue of their functions. Complement clauses appear to supply propositional arguments of a predicate, much as nominals supply referring arguments of a predicate. Relative clauses supply attributive modifiers of a noun much as (intersective) attributive adjectives supply attributive modifiers of a noun. As Givón (1990: 498) notes, when clauses take on “a prototypical nominal position (or function) ... within another clause” they are often nominalized.<sup>18</sup> Unlike English, which realizes finite/non-finite complement and finite/non-finite relative clauses in the same positions, *iPersian* sharply distinguishes the two types positionally, in both the verbal and the nominal domain.

*iPersian* is fundamentally a verb-final language with nominal arguments occurring almost exclusively before V. As many authors have noted, whereas *iPersian* non-finite complement clauses occur preverbally, like nouns, *iPersian* finite complements diverge in being uniformly postverbal. Compare (49)–(51), which illustrate a variety of *iPersian* construction types and where the clausal/clause-like complements are bracketed and where the verb is boldfaced to highlight its position relative to them.<sup>19</sup>

17. By ‘clausal projection XP’ we refer to any argument XP with propositional semantics – type <t> or <s,t> – or any attributive XP – type <e,t> – deriving from a phrase with propositional semantics, such as a relative clause.

18. For useful discussion of relative clause and clausal complement typology and nominal properties, see Lehmann (1986, 1988) and Schmidtke-Bode (2014).

19. N. Shafiei (p.c.) notes that presence/absence of the modal correlates with controller choice in the embedded verb. Without *bayæd* Ali is the controller; with *bayæd* control is ambiguous between Maryam and Ali.



- (49) a. **Finite control clause**  
*Mina Ali-ro qhane=kærd [ke (bayæd) be-re]*  
 Mina Ali-ACC persuaded that (should) SBJV-go.3SG  
 “Mina persuaded Ali that he/she should leave.”
- b. **Non-finite control clause**  
*Mina Ali-ro [be ræft.æn] qhane=kærd.*  
 Mina Ali-ACC to go.INF persuaded  
 “Mina persuaded Ali to leave.”
- (50) a. **Finite perception V complement clause**  
*Mina did [ke Ali ræft].*  
 Mina saw that Ali leave.PST.3SG  
 “Mina saw that Ali left.”
- b. **Non-finite percept V complement clause**  
*Mina [ræftæn=e Ali]-ro did.*  
 Mina go.INF=EZ Ali]-ACC see.PST.3SG  
 “Mina saw Ali leave’/Mina saw Ali’s leaving.”
- (51) a. **Finite clausal complement**  
*Mina færz=kærd [ke Ali gonahkar-e].*  
 Mina considered that Ali guilty-be.PRS.3SG  
 “Mina considered/assumed that Ali is/was guilty.”
- b. **Non-finite clause complement**  
*Mina [gonahkar budæn=e] Ali-ro færz=kærd.*  
 Mina guilty be.INF=EZ Ali-ACC considered  
 “Mina considered/assumed Ali to be guilty.”
- c. **Small clause complement**  
*Mina [Ali -ro gonahkar] færz=kærd.*  
 Mina Ali -ACC guilty considered  
 “Mina considered (assumed) Ali guilty.”

In each case the non-finite complements occur leftward of the verb, like nominal arguments, whereas the finite ones occur uniformly rightward. In Givón’s terms, then, iPersian nonfinite complements occupy “prototypical nominal positions” in VP and hence might be expected to show noun-like (iPersian) behavior. By contrast iPersian finite complements do NOT occupy prototypical nominal positions, and hence noun-like behavior is not expected.

A strikingly similar pattern holds in iPersian nominals. Samiian (1983, 1994) argues that the unmarked sequence of *Ezafe*-marked constituents is as in (52a), illustrated by (52b), where the outermost NP is a genitive:

- (52) a. [ N =EZ AP =EZ PP =EZ NP ]  
 b. [ [<sub>N</sub> *xune*] =ye [<sub>AP</sub> *kucik*] =e [<sub>PP</sub> *kenar=e dærya*] =ye [<sub>NP</sub> *Ali*]]  
       house =EZ       small =EZ       by =EZ   sea                = EZ Ali  
       “Ali’s small house by the sea”/  
       “The small house by the sea of Ali’s”

Interestingly, the genitive appears to mark a right edge in the NP domain analogous to the verb in the VP domain. That is, the pre-genitive domain includes uncontroversially nominal complements and modifiers of N, including non-finite propositional ones; these all bear *Ezafe*. By contrast, the post-genitive domain includes PPs with no *Ezafe*-marking (internally or externally) and finite propositional complements and modifiers of N.

To illustrate these points, Example (53a) shows the noun *qol* “promise” with a nonfinite propositional complement *amædæn(=e) be Paris* “coming to Paris”. *Ezafe* is required. (53b)–(54c) show that insertion of a genitive must occur at the right edge and not postnominally. Thus the nonfinite propositional complement must precede the genitive like other [+N] complements under (52a). (53d) shows that a goal PP complement *be Hæsæn* “to Hasan”, with no *Ezafe*-marking, must be added outside the genitive.

- (53) a. *qol* =e [*amædæn(=e) be Paris*]  
       promise =EZ come.INF=EZ to Paris  
       “the promise of coming to Paris”/“the promise to come to Paris”  
 b. *qol* =e [*amædæn(=e) be Paris*] =e [<sub>NP</sub> *Ali*]  
       promise =EZ come. INF=EZ to Paris =EZ Ali  
       “Ali’s promise of coming/to come to Paris”  
 c. \**qol* =e [<sub>NP</sub> *Ali*] =ye [*amædæn(=e) be Paris*]  
       promise -EZ Ali =EZ come.INF= EZ to Paris  
       “Ali’s promise of coming/to come to Paris”  
 d. *qol* -e [*amædæn(-e) be Paris*] -e [<sub>NP</sub> *Ali*] [*be Hæsæn*]  
       promise -EZ come.INF-EZ to Paris -EZ Ali to Hasan  
       “Ali’s promise to Hasan of coming/to come to Paris”

Compare now (54a)–(54d). Example (54a) shows the same noun *qol* “promise” with a finite propositional complement *ke miyad Paris* “that he’ll come to Paris”. *Ezafe* is now excluded. (54b)–(54c) show that insertion of a genitive must occur post-nominally, and not at the right edge. Thus the finite propositional complement must follow the genitive, outside the *Ezafe* domain. (54d) shows that the goal PP complement *be Hæsæn* “to Hasan”, with no *Ezafe*-marking, accompanies the finite clause outside the genitive.

- (54) a. *in qol* (\*=e) [*ke miyad Paris*]  
 this promise =EZ that come-to Paris  
 “the promise that he will come to Paris”
- b. *qol =e Ali* [*ke miyad Paris*]  
 promise =EZ Ali that come-to Paris  
 “Ali’s promise that he’ll come to Paris”
- c. \**in qol* [*ke miyad Paris*] =e [<sub>NP</sub> *Ali*]  
 this promise that come-to Paris =EZ Ali  
 “Ali’s promise that he’ll come to Paris”
- d. *qol =e Ali* [*be Hæsæn*] [*ke miyad Paris*]  
 promise =EZ Ali to Hasan that come-to Paris  
 “Ali’s promise to Hasan that he’ll come to Paris”

An identical pattern is observed with relative clauses. As noted above, reduced – i.e., non-finite – relative clauses require *Ezafe* (55a). Insertion of a genitive must occur at the right edge of the noun phrase (55b), not postnominally (55c). The nonfinite propositional modifier thus patterns like other [+N] attributives under (52a). (55d) shows that the PP modifier *bi Hæsæn* “without Hasan”, with no *Ezafe*-marking, must be added outside the genitive, outside the *Ezafe* domain.

- (55) a. *æks* \*(=e) [*čap=šode dær ruzname*]  
 photo =EZ published in newspaper  
 “the photo published in the newspaper”
- b. *æks =e* [*čap=šode dær ruzname*] =ye [<sub>NP</sub> *Ali*]  
 photo =EZ published in newspaper =EZ Ali  
 “Ali’s photo published in the newspaper”
- c. \**æks =e* [<sub>NP</sub> *Ali*] =ye [*čap=šode dær ruzname*]  
 photo =EZ Ali =EZ published in newspaper  
 “Ali’s photo published in the newspaper”
- d. *æks =e* [*čap=šode dær ruzname*] =ye [<sub>NP</sub> *Ali*] [*bi Hæsæn*]  
 photo =EZ published in newspaper =EZ Ali without Hasan  
 “Ali’s photo without Hasan published in the newspaper”

Compare now (56a)–(56d). Example (56a) shows that finite relative clauses reject *Ezafe*. (56b) and (56c) show that insertion of a genitive must occur postnominally, and not at the right edge. Thus the finite relative must occur after the genitive, outside the *Ezafe* domain. (56d) shows that the PP modifier *bi Hæsæn* “without Hasan”, with no *Ezafe*-marking, must accompany the finite relative outside the *Ezafe* domain.

- (56) a. *æks* \*(=e) [*ke čap=šode-bud dær ruzname*]  
 photo =EZ that published-be.PST in newspaper  
 “the photo that had been published in the newspaper”

- b. *æks* =*e* [<sub>NP</sub> *Ali*] [*ke* *čap*=šode-*bud* *dær* *ruzname*]  
 photo =EZ *Ali* that published-be.PST in newspaper  
 “Ali’s photo that had been published in the newspaper”
- c. \**æks* [*ke* *čap*=šode-*bud* *dær* *ruzname*] =*ye* [<sub>NP</sub> *Ali*]  
 photo that published-be.PST in newspaper =EZ *Ali*  
 “Ali’s photo that had been published in the newspaper”
- d. *æks* =*e* [<sub>NP</sub> *Ali*] [*bi* *Hæsæn*] [*ke* *čap*=šode-*bud*  
 photo =EZ *Ali* without Hasan that published-be.PST  
*dar* *ruzname*]  
 in paper  
 “Ali’s photo without Hasan that had been published in the paper”

iPersian thus appears to distinguish nominal versus non-nominal elements positionally within VP and NP, and in similar ways. Within VP, nominal arguments are predominantly preverbal whereas non-nominal elements can, and in some cases must, appear postverbally. Non-finite propositional complements pattern like nominals in occurring preverbally whereas finite complements are always postverbal. Within NP, nominal ([+N]) elements are pre-genitival whereas non-nominal elements are post-genitival. Non-finite complement and relative clauses pattern like [+N] elements in requiring *Ezafe* and in occurring before the genitive, whereas finite complement and relative clauses occur uniformly after the genitive. Prediction 1 regarding *Ezafe* with clausal projections (47a) is thus supported. iPersian *Ezafe* appears ONLY before clausal elements showing the external positional distribution of [+N] elements.

### 3.4.1.2 Prediction 1: iPersian vs. Sorani and Kurmanji

We noted above that whereas iPersian forbids *Ezafe* before finite relative clauses (FRCs), both Central Kurdish (Sorani) and Northern Kurdish (Kurmanji) require it (57). Sorani and Kurmanji FRCs thus resemble iPersian RRCs in requiring *Ezafe*.

- (57) a. in *dastan* (\*=*e*) [*ke* *be* *mæn* *goft*] iPersian FRC  
 this story =EZ that to me say.PST.3SG  
 “The story that he told me”
- b. *chirok*-*æka*=*y* [(*kæ*) *æw* *bæ* *mn*-*I* *ku*] (Sorani FRC)  
 story-DEF-EZ that he to me-CL.3SG told  
 “The story that he told me” (Abdollahnejad p.c.)
- c. *çirok*=*a* [*ku* *wî* *ji* *min* *re* *got*] (Kurmanji FRC) (Songül Gündoğdu p.c.)  
 story=EZ.FEM that 3S.OBL ADP 1S.OBL ADP say.PST.3SG  
 “The story that he told me”

Compare now the situation with finite clausal complements (FCCs) of nouns. In all three languages, *Ezafe* is forbidden (58):

- (58) a. *in omid* (\*=e) [*ke Shah æz Iran xahæd.ræft*] (iPersian FCC)  
 this hope =EZ that Shah from Iran will-go  
 “the hope that the Shah will leave Iran.”
- b. *æw hiwa=yæ-(y)* [*kæ Sha læ Eran(=e) dæ-rw-a*] (Sorani FCC)  
 this hope=EZ-? That Shah from Iran=EZ will-go-3SG  
 “this hope that the Shah will leave Iran” (Elias Abdollahnejad p.c.)
- c. *ev hêvî=ya* [*ku Shah ji Iran-ê derkev-e*]  
 this hope=EZ.FEM that Shah from Iran-OBL SUBJ.go out-3SG  
 “the hope that the Shah will leave Iran.” (Kurmanji FRC)  
 (Songül Gündoğdu p.c.)

The case-marking analysis of *Ezafe* makes straightforward predictions about the source of these cross-linguistic patterns. If *Ezafe* is required before [+N] categories and blocked before [-N] categories, then the distribution in (57)–(58) must reflect varying ‘nominality’ in the relevant clause types. Specifically,

Predictions: iPersian FRCs are non-nominal ([-N])  
 Kurmanji/Sorani FRCs and FCCs and are nominal ([+N])

Under the widely held view that clauses are CPs projected from their complementizers I, these become predictions about the featural composition of the various C heads:

Predictions: In iPersian FRCs, C is non-nominal  
 (i.e., a non-nominal complementizer)  
 In Kurmanji and Sorani FRCs and FCCs, C is nominal

We thus derive the following claims about the complementizer inventories of iPersian, Sorani and Kurmanji:

- iPersian has a [-N] complementizer *ke*, occurring in FRC and FCCs
- Sorani has a [+N] relative pronoun *ka*, occurring in FRCs.
- Sorani has a [+N] complementizer *ka*, occurring in FCCs.
- Kurmanji has a [+N] relative pronoun *ku*, occurring in FRCs.
- Kurmanji has a [+N] complementizer *ku*, occurring in FCCs.

In brief, then, iPersian is predicted to have a uniformly non-nominal complementizer inventory whereas Kurmanji and Sorani have uniformly nominal complementizer inventories.

The situation hypothesized here for Kurmanji and Sorani, with homophonous complementizer-relative pronoun pairs (*ka-ka/ku-ku*, respectively), is familiar from other languages. As Manzini (2010) observes, the pattern is widely attested in Romance. In (59) from Italian, the CP element *che* appears as a sentential

complementizer in (59a), as a relative pronoun in (59b), as a *wh*-phrase in (59c), and as a *wh*-determiner in (59d).

- (59) a. *So che fai questo*  
 Know.1SG **that** do.2SG this  
 “I know that you do this.”
- b. *Il lavoro che fai è noto*  
 the work **that** do.2SG is well-known  
 “the work that you do is well-known”
- c. *Che fai?*  
**What** do.2SG  
 “What are you doing?”
- d. *Che lavoro fai?*  
**Which** job do.2SG  
 “Which job do you do?”

Note that in at least (59c)–(59d), *che* has uncontroversial nominal character, which Manzini (2010) argues to be the general situation with Romance complementizers.

We cannot pursue the full consequences of these predictions here. But some preliminary data suggest they may be on the right track. Cross-linguistically, resumptive pronouns occur in relative clauses introduced by a complementizer, but not in ones introduced by a relative pronoun (McCloskey 2002; Lavine 2003; Merchant 2004; Citko 2004). In other words, relative pronouns and resumptive pronouns are mutually exclusive (Downing 1978). Interestingly, iPersian and Kurmanji FRCs appear to differ in this respect. As shown in (60) (from Aghaei 2006), iPersian FRCs permit resumptive pronouns in non-subject positions (boldfaced):

- (60) a. *doxtær-I [ke mæn (un-o) dus-eš dar-æm] vared-e kelas šod*  
 girl-INDF that I (**her-ACC**) friend-**her** have-1SG entry-EZ class did  
 “The girl whom I like (**her**) came into the class.”
- b. *šæhr-I [ke Ali (dar un) zendegi=mi-kon-e] æz inja dur-e*  
 city-INDF that Ali (**in that**) life=DUR-do-3SG. From here far-is  
 “The city where Ali lives (**in there**) is far away from here.”

By contrast, Kurmanji FRCs do not permit resumptive pronouns in non-subject positions. In direct object positions resumptive pronouns are simply ungrammatical (61):<sup>20</sup>

20. We are grateful to Songül Gündoğdu (p.c.) for the data in (61)–(63) and their discussion.

- (61) a. *keçik=a* [ku min (\*wê) doh dît] zehf rind bû  
 Girl=EZ.F that 1S.OBL **her** yesterday see.PST.3SG very pretty was  
 “the girl whom I saw (**\*her**) yesterday was very beautiful.”  
 b. *mal=a* [ku ez (\*wê) çû-m] zehf xweş bû  
 house=EZ.F that 1S.DIR **it** go.PST.1SG very nice be.PST.3SG  
 “the house that I went to (**\*it**) was very nice.”

In P-object position, Kurmanji must appeal to a ‘contracted adposition’ strategy that suppresses the pronoun. Thus in (62a) the contracted adposition *jê* appears in place of the full PP with pronoun *ji wê* (“to her”), which is ungrammatical (62b). Similarly in (63a), the contracted *lê* appears in place of *li wî* (“in there”), which is again ungrammatical (63b).

- (62) a. *keçik=a* [ku min *jê* ra gul şand]  
 Girl=EZ.F that 1S.OBL **ADP.3S.OBL** Part. rose.DIR send.PST.3SG  
*çû Stenbol-ê*  
 go Istanbul-OBL  
 “The girl whom I sent roses [to her] went to Istanbul”  
 b. \**keçika* [ku min *ji wê* ra gul şand] çû Stenbolê.  
 (63) a. *şehr=a* [ku ew *lê* di-jî] ji vir dûr e  
 City=EZ.F that 3S.DIR **ADP.3S.OBL** PROG-live.PRS.3S **ADP** here far is  
 “The city where s/he lives (in there) is far away from here.”  
 b. \**şehra* ku ew *li wî* dijî ji vir dûr e.

Given the generalization above, the possibility of resumptive pronouns in iPersian FRCs suggests that *ke* is a simple complementizer. By contrast, the impossibility of resumptive pronouns in Kurmanji is explained if *ku* is a relative pronoun. These results thus provide tentative support for our predictions.

### 3.4.1.3 Prediction 2: *Ezafe* recursion with RCs

Whereas Prediction 1 concerns the external character of a complement or relative clause – the status of the larger projection as [ $\pm$ N], Prediction 2 concerns elements internal to the clause – their character as (non-)nominal and hence potential hosts for an adnominal clitic like *Ezafe*.

External vs. internal nominality bears on the possibility of *Ezafe* recursion relative clauses. We noted that iPersian participial relatives behave like other [+N] items in so far as they bear *Ezafe* and occur leftward of a genitive. This accords with the general, widely-observed nominal character of participial clauses (Krause 2001). We also noted in (29a), (29b) (repeated below as (64)) that iPersian participial relatives show recursion with *Ezafe*. Under the analysis of *Ezafe* as a nominal clitic (or nominal morphology), this possibility requires [+N] status for the clause-final

participle (*bargašte*). Again this accords with widely noted generalizations regarding the adjectival ([+N]) status of lexical participles.

- (64) a. *jævan=e* [æz Swis bærgæšt-e]=ye [estexdam  
young man=EZ from Swiss return.PST-PTCP=EZ employment  
šod-e dær vezaræt=e færhæng] (iPersian RRC)  
get.PST-PTCP in ministry-EZ education  
“the young man back from Switzerland employed by the Ministry of  
Education”
- b. *dust=e* [æz Swis bærgæšte]=ye [jævan-i ke molaqat-kærd-i]  
friend=EZ from Swiss returned=EZ youth-IND that meet-did-2sg  
“the recently returned friend from Switzerland of the young man that you  
met” (iPersian RRC)

Thus the combined external-internal nominal character of iPersian participial relatives correctly predicts the possibility of *Ezafe* recursion.

Compare now the behavior of Northern Kurdish (Kurmanji) and Middle Kurdish (Sorani). Whereas both Kurdish variants exhibit *Ezafe* before a finite relative clause (65), only Kurmanji allows *Ezafe* recursion (65a); Sorani rejects it (65b).<sup>21</sup>

- (65) a. *çîrok=a* [ku wî ji min re got]  
Story=EZ.FEM that 3s.OBL ADP 1s.OBL ADP say.PST.3s  
“The story that he told me” (Kurmanji, Songül Gündoğdu p.c.)
- b. *chîrok-ækæ=y* [(kæ) æw bæ mn-i kut]  
story-DEF=EZ that he to me-CL.3SG told  
“The story that he told me” (Sorani FRC, Elias Abdollahnejad p.c.)

21. Interestingly, although Kurmanji and Sorani exhibit participial modifiers with *Ezafe*, participial relative clauses are apparently unavailable as opposed to full FRCs. These points are illustrated by (i)–(ii) from Kurmanji (Songül Gündoğdu p.c.):

- (i) a. *birîn=a* *dermankir-î*  
Wound=EZ.F treat-PRT  
“The wound treated”
- b. *nan=ê* *may-î*  
bread= EZ.M stay-PRT  
“The bread being left over”
- (ii) a. \**birîn=a* [bi destê Betul *dermankir-î*]  
Wound=EZ.F by Betul treat-PRT  
Intended: “The wound treated by Betul”
- b. *birîn=a* [ku bi destê Betul *hatî-ye dermankirin*]  
wound=EZ.F that by Betul come.PST-3SG to treat]  
“The wound treated by Betul”



- (66) a. *çîrok=a [ku wî ji min re got] ya [ku*  
 Story=EZ.F that 3S.OBL ADP 1S OBL ADP say.PST.3S. EZ.F that ADP  
*di rojnamê da derket ]*  
 newspaper.OBL PART come out.PST.3SG  
 “The story that he told me that was published in the newspaper”<sup>22</sup>  
 (Kurmanji, Songül Gündoğdu p.c.)
- b. *chirok-aka=y [(ka) aw ba amn-i kut] (\*=y) [(ka) la rozhnama*  
 story-DEF=EZ that he to me-CL.3SG told =EZ that in newspaper  
*da blaw bo-ta-wa ]*  
 in publish has-1SG-been  
 “The story that he told me that has been published in the paper.”  
 (Sorani, Elias Abdollahnejad p.c.)

Thus whereas both Kurmanji and Sorani finite relative clauses must be externally nominal under the case-marking analysis (Prediction 1), it appears Kurmanji must also be internally nominal – i.e., the apparent finite verb stem *got* “say.PST.3s” must be underlyingly nominal, despite surface appearances. Otherwise *ya* would not be attaching to a nominal stem, contra assumptions (Prediction 2).<sup>23</sup>

22. Songül Gündoğdu reports that Kurmanji speakers accept (29c); but they regard (i), where *Ezafê* attaches to an overt pronominal, as more natural:

- (i) *çîrok=a [ku wî ji min re got] ew-a [ku di*  
 Story=EZ.F that 3S.OBL ADP 1S OBL ADP say.PST.3S. EZ.F that ADP  
*rojnamê da derket]*  
 newspaper.OBL PART out.PST.3SG  
 “The story that he told me that was published in the newspaper”

The status of the pronominal element in *ew-a* is unclear to us. Gündoğdu (p.c.) suggests it might be an instance of so-called ‘demonstrative/anaphoric *Ezafê*’ (Haig 2011). If so the gloss of (i) is actually closer to ‘The story that he told me, the one that was published in the newspaper’. If ‘one’ takes ‘story-published-in-the-newspaper’ as its antecedent, this will be equivalent to the standard interpretation of recursive RCs as expressing successive intersection.

23. An alternative proposal discussed Larson et al. (2019) is that *Ezafê* is reanalyzed in Kurmanji from being nominal morphology to a syntactic clitic counterpart to the English common genitive ‘s, which cliticizes freely onto a [+N] PHRASE to its left:

- (i) a. *[Fred] ’s opinion about the English genitive is different from mine.*  
 b. *[The man on the Clapham omnibus] ’s opinion about the English genitive is poorly thought out.*  
 c. *[Every linguist I know] ’s opinion about the English genitive involves functional categories.*  
 d. *[That young hotshot who was recently hired at Princeton that I was just telling you about] ’s opinion about the English genitive is simply wrong.*  
 e. *Even [that colleague who shares an office with you] ’s opinion about the English genitive is not to be trusted.*  
 (from Anderson 2013)

Although we do not yet possess clear evidence for the correctness of these conjectures, we do note that Kurmanji is the Kurdish variant occurring in closest geographical proximity to Turkish, a language in which both complement and relative clauses are well-known to display internal nominalization (Göksel & Kerslake 2005). It seems to us at least plausible that Kurmanji's behavior might represent an areal effect. We must leave this possibility for future investigation.

### 3.4.2 Predictions of the case-marking analysis: PPs

The case-marking analysis also makes clear predictions with regard to iPersian NP-PP modifier structures and the distribution of *Ezafe*. We saw that iPersian prepositional forms divide into three classes:

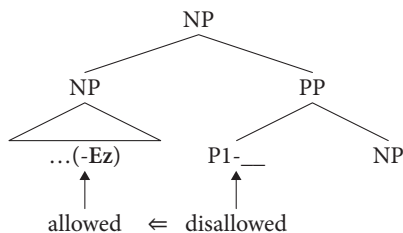
Class 1: P's that **DISALLOW** *Ezafe* between themselves and their complement.

Class 2: P's that **ALLOW** *Ezafe* between themselves and their complement.

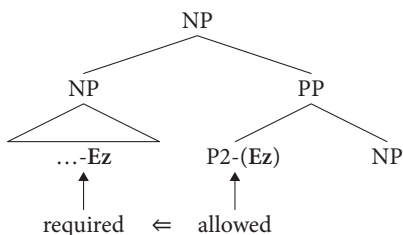
Class 3: P's that **REQUIRE** *Ezafe* between themselves and their complement.

We furthermore noted that P-class appears to condition occurrence of *Ezafe* on NP in NP-PP modifier structures. The pattern was as in (35) (repeated below):

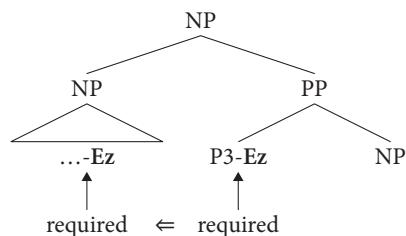
(35) a. **-EZ and P1's**



b. **-EZ and P2's**



c. **-EZ and P3's**



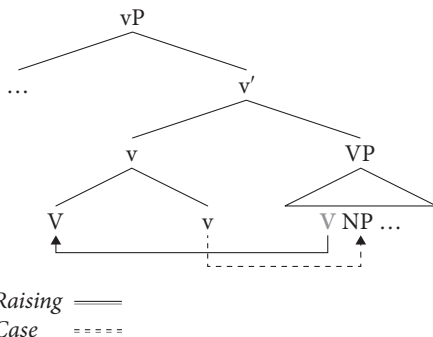
Under the case-marking analysis, the distribution in (35) should reflect the nominal nature of PP. Specifically, it should be the case that, when headed by a P2 or a P3, PP is unambiguously nominal in character and hence requires *Ezafe* before it. But when headed by a P1, PP must be somehow ‘optionally nominal’, allowing *Ezafe* to be present or absent.

Larson & Samiian (2018) argue that the case-marking analysis can accommodate these facts through an elaboration of ideas by Jackendoff (1973, 1977) and Svenonius (2003) on the relation between VP and PP structure. Jackendoff (1973) establishes a basic parallelism in the complementation of V and P, with the verbal patterns in (67a)–(67d) matching the prepositional patterns in (67a')–(67d')

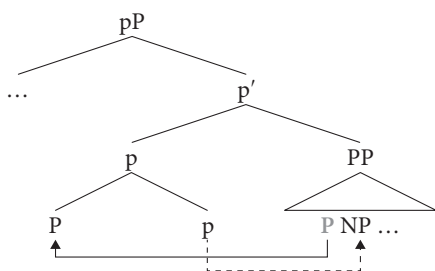
(67) Verbal complementation	Prepositional complementation
a. [ <sub>VP</sub> V] <i>laugh, cough, run, fall, etc.</i>	a'. [ <sub>PP</sub> P] <sup>24</sup> <i>in(side), down, out, over, etc.</i>
b. [ <sub>VP</sub> V NP] <i>hit, kiss, see, etc.</i>	b'. [ <sub>PP</sub> P NP] <i>in(side), down, out, over, etc.</i>
c. [ <sub>VP</sub> V PP] <i>dash, emerge,, reply, etc.</i>	c'. [ <sub>PP</sub> P PP] <i>into, down, from, up, etc.</i>
d. [ <sub>VP</sub> V NP PP] <i>give, send, put, etc.</i>	d'. [ <sub>PP</sub> P NP PP] <i>into, down, from, to, in, etc.</i>

In recent work these parallels have been developed further to include recognition of a functional head *p*, which assigns a ‘figure’/‘locatum’ role in PP and case to an object (van Riemsdijk 1990; Svenonius 2003), much as *v* head assigns the agent role in VP and case to an object (Chomsky 1995). In both structures, the lexical head raises to the corresponding functional head (68):

(68) a. VP Structure

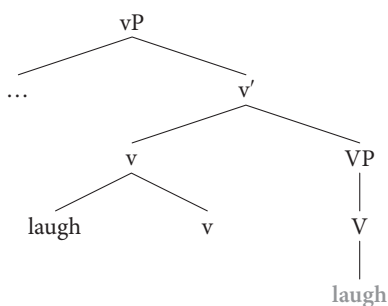


24. See Klima (1965) and Emonds (1976) for original arguments for these forms as ‘intransitive prepositions’.

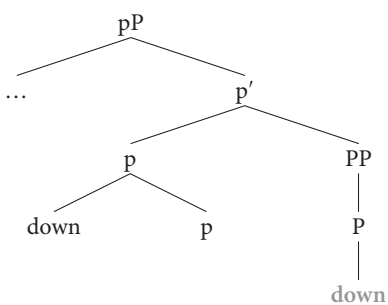
b. *PP Structure*

Examples (69a)–(69c) display comparable vP/pP structures; note that head raising is non-string vacuous in (69c)/(69c').

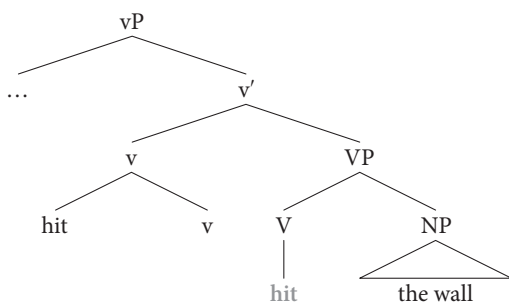
(69) a.

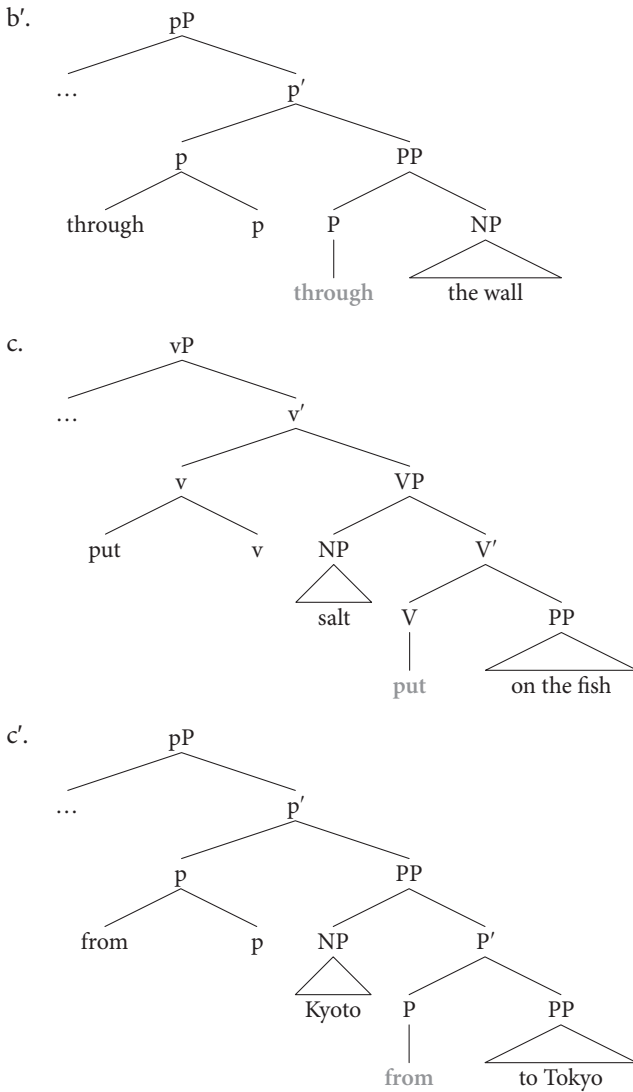


a'.



b.





Larson & Samiian (2018) propose to capture *Ezafé* distribution with iPersian PPs by drawing an additional vP/pP parallelism in the domain of nominalization. Consider the internal form and external behavior of the boldfaced phrases in (70)

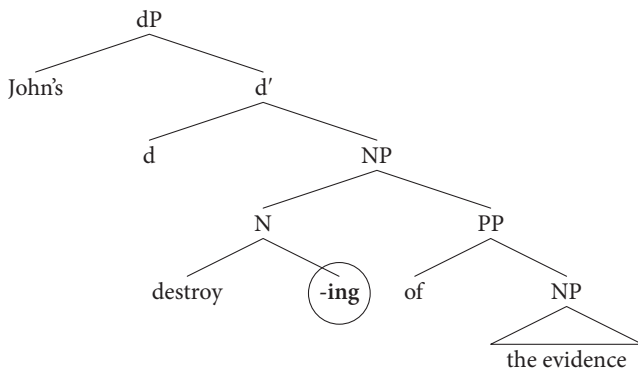
- (70) a. [<sub>VP</sub> V NP]      John will **destroy the evidence**      *of*-forbidden
- b.i. [<sub>NP</sub> V-*ing* NP]      John's **destroying the evidence**
- ii. [<sub>NP</sub> V-*ing* of NP]      John's **destroying of the evidence**      } *of*-optional
- c. [<sub>NP</sub> N *of* NP]      John's **destruction of the evidence**      *of*-required



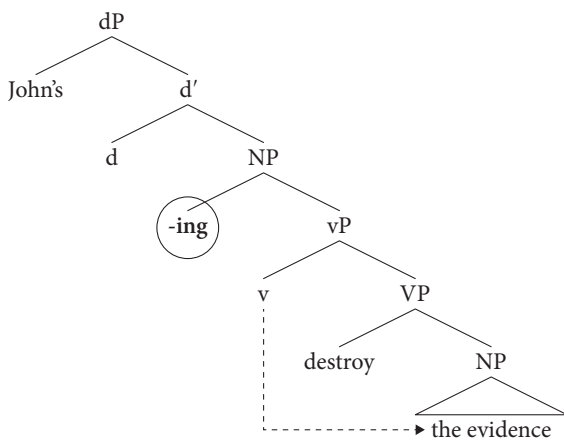
- c. *in/un pošt=e mašin*  
**this/that** behind=EZ car  
 “This/that back area of the car”
- (73) a. *un zir-a=ye miz* (= K&B (45a), (45b), (45d))  
 that under-PL=EZ table  
 “Those under spaces of the table”  
 b. *un vāsæt-a=ye otaḡh*  
 that middle-PL=EZ room  
 “Those middle parts of the room”  
 c. *in pošt-a=ye xune*  
 this behind-PL=EZ house  
 “These back areas of the house”
- (74) a. *be zir=e miz* (= K&B (46a))  
**to** under=EZ table  
 “Under (directional) the table”  
 b. *zir=e kāsif=e miz* (= K&B (47))  
 under=EZ **dirty=EZ** table  
 “The dirty underspace of the table”

Larson & Samiian’s (2018) account of iPersian P2s extends Jackendoff’s (1977) analysis of nominal vs. verbal behavior in gerunds. Jackendoff proposes that in nominal gerunds, a nominalizing morpheme *-ing* attaches to the lexical V, converting it to N and determining its projection as NP. By contrast, in verbal gerunds, the nominalizer attaches to the larger VP phrase, converting it to an NP, but leaving its internal verbal structure intact. We update Jackendoff’s proposals for gerunds slightly in (75a)–(75b) below. Note that *-ing*’s positioning above vP in (75b) allows v to assign accusative case to the object. Structures for the corresponding derived nominal and simple vP are given in (75c)–(75d), respectively.

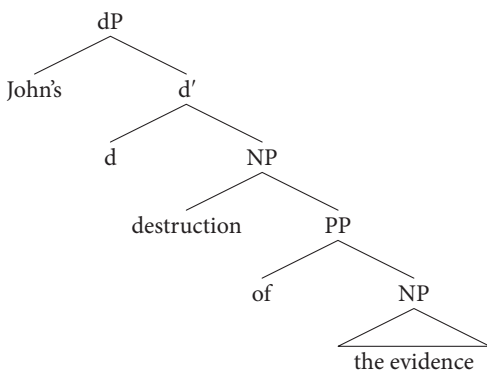
- (75) a. **Nominal gerund** (nominalized V)



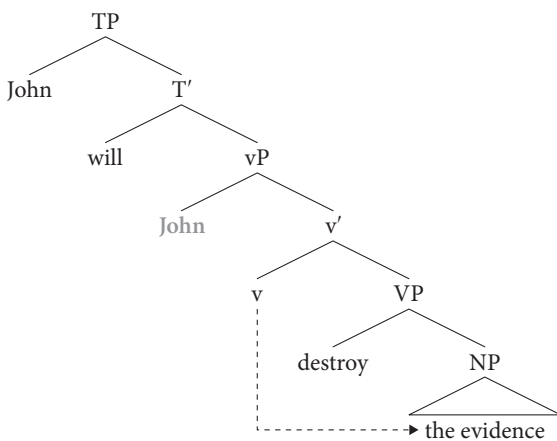
## b. Verbal gerund (nominalized vP)



## c. Derived nominal (deverbal N)



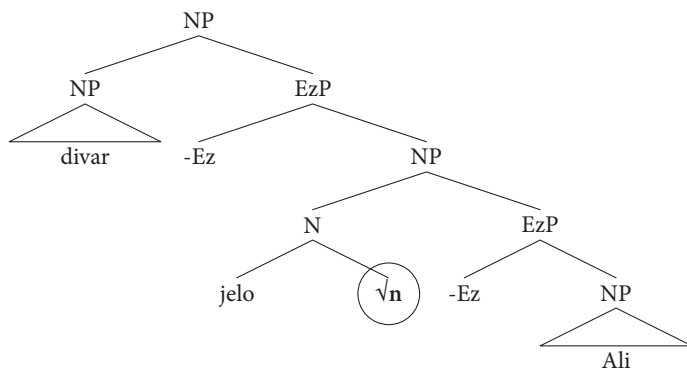
## d. Simple vP





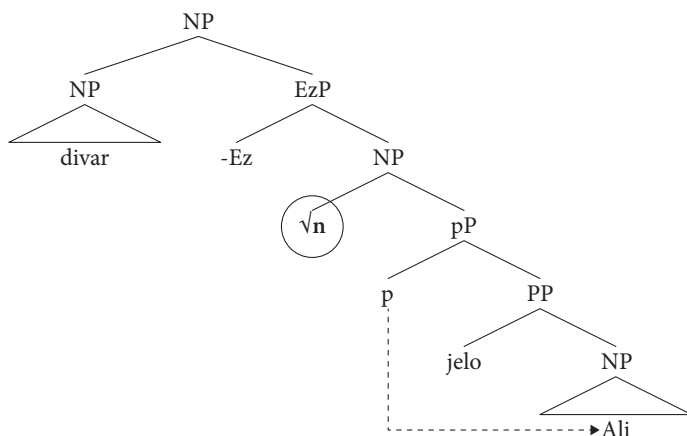
Larson & Samiian (2018) propose a fully analogous account of iPersian Class 2 forms. Specifically, they propose that when a Class 2 P appears WITH a following *Ezafe*, a nominalizing morpheme  $\sqrt{n}$  has attached to P, converting it to N and determining its projection as NP (76a). This form is the prepositional counterpart of a nominal gerund in containing a nominalized HEAD (cf. (75a)). By contrast, when a Class 2 P appears WITHOUT a following *Ezafe*, the nominalizer has attached to the larger pP, converting it to an NP, but leaving its internal prepositional structure intact (76b). This form is the prepositional counterpart of a verbal gerund in containing a nominalized PHRASE (cf. (75b)). Here again,  $\sqrt{n}$ 's position above pP in (76b) allows p to assign accusative case to the object. Structures for the corresponding P3s and P1s are given in (76c)–(76d), respectively. The parallelism to verbal nominalizations is evidently quite close.<sup>25</sup>

(76) a. Class 2 (nominalized P)

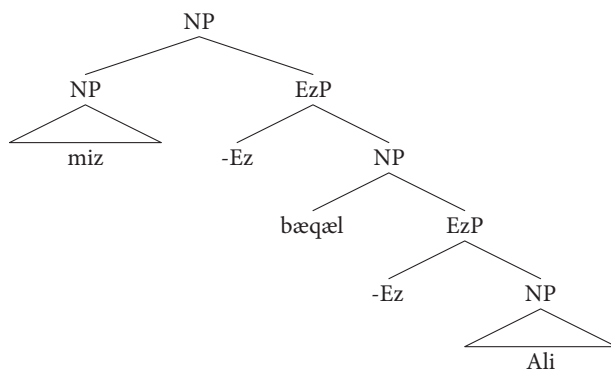


25. A technical question arises as to how co-occurrence between Class 2 forms and the nominalizer  $\sqrt{n}$  is ensured. Larson & Samiian (2018) propose to extend the account of formal features in Pesetsky & Torrego (2007) to category features. More precisely, they propose that nominalization involves separate instances of a nominal feature [N], one interpretable (iN) and one valued ([Nval]), which must enter an agreement relation for legibility at the LF-PF interfaces. Class 2 forms are proposed to bear a [Nval] feature lexically, which then requires a c-commanding  $\sqrt{n}$  bearing [iN] to come into agreement with it. It is the interpretable instance of [N] that determines semantic scope of nominalization in the sense of Jackendoff (1977).

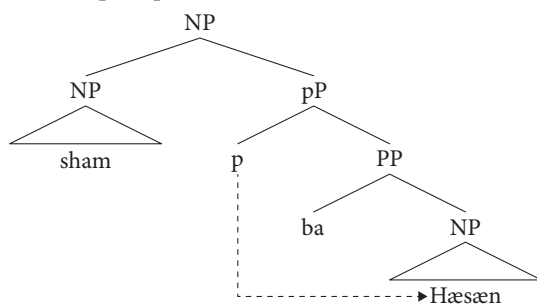
## b. Class 2 (nominalized pP)



## c. Class 3 (de-prepositional N)



## d. Class 1 (pure pP)



Finally, Larson & Samiian (2018) analyze the absence of *Ezafe* after P1s and its optionality on the phrase that P1s head as analogous to what one sees in English with (77).

- (77) a. *John's destroying (of) the evidence (was illegal).*  
 b. *John's borrowing (of) the tools (was frowned on).*  
 c. *John's hearing (\*of) the noise (was unexpected).*  
 d. *John's knowing (\*of) French (was not taken for granted).*  
 e. *John's loving (\*of) chocolate (was a drawback).*

It is well known that whereas virtually any verb in English can occur in a verbal gerund, occurrence in a nominal gerund is more restricted and constrained by the verbal semantics. Specifically, whereas action verbs readily form nominal gerunds (77a)–(77b), stative predicates including verbs of perception or mental attitude do not (77c)–(77e).<sup>26</sup> This pattern is natural under Jackendoff's scopal analysis; we expect lexical constraints to exert themselves when nominalization applies to the lexical stem, but not when it applies to the phrasal projection.

Larson & Samiiian (2018) analyze the *Ezafe* facts with P1s in a parallel way. The proposal is that whereas P1s reject nominalization as a matter of their lexical semantics, the pP phrase they project more readily accepts nominalization since lexical constraints are not in play. Thus whereas (78a) is excluded, (78b) is acceptable in various instances (cf. (35c)).<sup>27</sup>

---

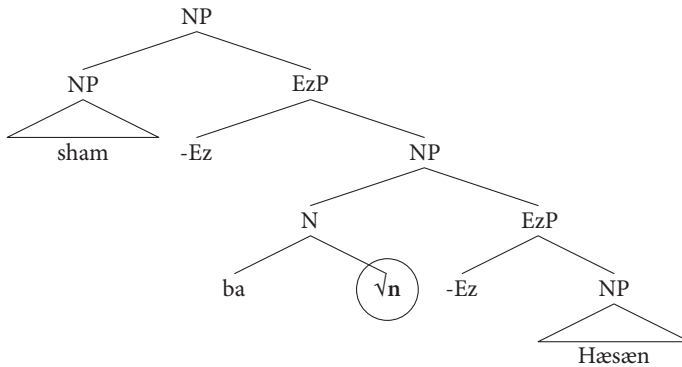
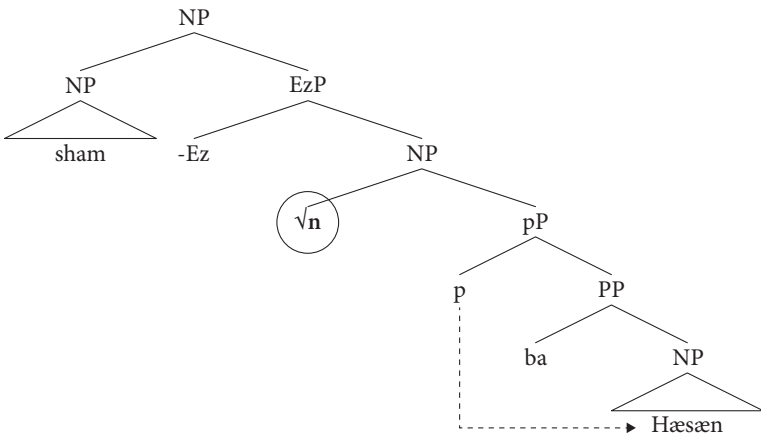
26. Similar constraints are found in progressives (ia), (ib), suggesting that gerund and progressive *-ing* are related.

- (i) a. \*John is knowing French.  
 b. \*John is loving chocolate. (must mean 'loving eating')  
 c. \*John is believing that climate change has occurred.

27. A corpus study was conducted by Nazila Shafiei of PPs headed by 6 P1 prepositions (*dar* "in/inside", *bar* "on/onto", *be* "to/toward", *az* "from", *ta* "until/to" and *ba* "with") of the first fifty thousand lines of the Bijankhan corpus. PPs were categorized as not allowing, optionally allowing and requiring a preceding *Ezafe*. A total of 126 occurrences were recorded. The majority of the cases did not allow an external *Ezafe* (i); some cases allowed an optional *Ezafe* (ii); there was only one instance of a required *Ezafe* (iii)

- (i) *tæhsil* (\*=*e*) [<sub>PP</sub> *dær* [*reshte=ye honær-ha=ye ziba*](\*)=*e*] [<sub>PP</sub> *dær daneshgah*]  
 education \*=EZ in field=EZ art-PL=EZ fine \*=EZ at university  
 'education in the field of fine arts at the university'
- (ii) [<sub>NP</sub> *eshq* (=e) [<sub>PP</sub> *be zendegi*]]  
 love =EZ for life  
 "love of life"
- (iii) [<sub>NP</sub> *goruh* \*(=e) [<sub>PP</sub> *dær shæhr*]]  
 group =EZ in city  
 "the group in the city",

(78) a. \*Class 1 (nominalized P)

b. <sup>OK</sup> Class 1 (nominalized pP)

In summary, the case marking analysis predicts that occurrence of *Ezafe* internally to PP in iPersian should be a matter of the ‘nominality’ of the P head. And occurrence of *Ezafe* externally to PP should be a matter of the nominality of the phrase that P projects. This prediction is transparently correct in the case of P3s, as argued by Karimi & Brame (1986, 2012); here the head and phrase are both N. The P2 and P1 classes, which have previously escaped systematic treatment, can be assimilated into this picture in an enlightening way by extending Jackendoff’s (1973, 1977) proposals regarding the structure of PP and scopal nominalization to the iPersian prepositional system.

### 3.4.3 Predictions of the case-marking analysis: Cross linguistic variation

In addition to the predictions the case-marking analysis makes for familiar *Ezafe* phenomena from iPersian and Kurdish, we wish to briefly draw attention to its relevance for a wider data set, including the so-called ‘doubled’ or ‘strengthened’

*Ezafe* construction in Zazaki, and the ‘Reverse *Ezafe*’ construction observed in the Caspian languages (Gilaki, Mazanderani, Talyshi) and possibly in Balochi.

### 3.4.3.1 Zazaki ‘doubled *Ezafe*’

Zazaki exhibits *Ezafe* in the same structural contexts as other Iranian languages, but Zazaki *Ezafe* morphology is especially complex. As discussed in Todd (1985), from which the examples below are drawn, the form of the *Ezafe* in (79) encodes gender (masculine vs. feminine), number (singular vs. plural), and whether the relation between N and its complement is descriptive/adjectival vs. genitival/possessive:

- (79) a. *pir'tok=o find*  
Book=EZ good  
“good book”
- b. *suk=a gird-i*  
city=EZ large-fem  
“large city”
- c. *ban=e min*  
house=EZ me(OBL)  
“my house”
- d. *ling=a min*  
foot=EZ me (OBL)  
“my foot”
- e. *sa=y wes-i*  
apple=EZ good-PL  
“good apples”
- f. *ling=e min*  
feet=EZ me(OBL)  
“my feet”

A unique feature of Zazaki is its so-called ‘doubled’ or ‘strengthened’ *Ezafe*. When a phrase containing *Ezafe* is embedded in a larger *Ezafe* construction, the embedded *Ezafe* morpheme (EZ) shows a special form, becoming *de* or *da* (DEZ) depending on gender and/or number. This situation is schematized in (80) and illustrated with examples in (81):

- (80) a. [HEAD=EZ [HEAD =**de** MOD]] (masculine or plural)  
b. [HEAD=EZ [HEAD =**da** MOD]] (feminine)
- (81) a. *kutik=e [əmiryan=**de** ma]*  
Dog=EZ neighbor(OBL)=DEZ us  
“our neighbor’s dog”
- b. *aqil=e [mar'dim=**de** pil-i]*  
wisdom=EZ people=DEZ older-PL  
“the wisdom of older people”



- c. *ma=y [mar=da ay]*  
 mom=EZ mom(OBL)=DEZ her  
 “her mother’s mother”

Interestingly, *Zazaki Ezafe* exhibits the very same shape change when a phrase containing *Ezafe* is the object of an oblique postposition, as shown in (82):

- (82) a. [HEAD=**de/da** MOD ] P  
 b. [*embaz=de xwi*] -*re*  
 friend=DEZ own -to  
 “to his friend”  
 c. [*mar-da to*] *fa*  
 mom(OBL)=DEZ you(OBL) from  
 “from your mother”

Thus *Ezafe* and oblique prepositions pattern together in their effect on a subordinate *Ezafe*.

Larson (2018) argues that *Zazaki* doubled *Ezafe* can be seen as part of a broader pattern of phenomena involving the case that is checked on DPs by external elements like T, v and p (83a), and a genitive case checked within DP (83b):

- (83) a. T/v/p DP  
 DP-external Case  
  
 b. [<sub>DP</sub> ... D ... NP ... AP ...] DP-internal Case  


Consider first the pair in (84) below, described by Babby (1987, 1988). As Babby observes, Russian quantified nominals exhibit an alternation in internal case marking, depending on their external environment. When the nominal is in a position of oblique case marking, the D, its modifiers, and the head of NP all inflect homogeneously for the externally assigned oblique case (84a). However, when the nominal is in a position of structural Case marking, only the D head is inflected for the external structural Case. The modifiers and the head of NP all inflect with genitive case, which Babby identifies as an internal case assigned by D (84b):

- (84) a. *a [pjat’ju bolšimi butylkami vina]* INST  
 with five.INST big.INST.PL bottle.INST.PL wine.GEN  
 “with five big bottles of wine”  
 b. *vypil [pjat’ bolšix butylok vina]* ACC  
 drank five.ACC big.GEN.PL bottle.GEN.PL wine.GEN  
 “drank five big bottles of wine” (Babby 1988: 289)

The examples in (85) show that alternative case patterns are not possible. It is not possible to inflect only D for external case in a position of oblique case marking (85a). Likewise it is not possible to inflect the internal elements of DP for structural Case in a position of structural Case-marking; DP-internal genitive case must appear (85b):

- (85) a. \*a [pjat'ju bolšix butylok vina ] INST  
 with five.INST big.GEN.PL bottle.GEN.PL wine.GEN  
 “with five big bottles of wine”  
 b. \*vypil [pjat' bolšie butylki vina ] ACC  
 drank five.ACC big.ACC.PL bottle.ACC.PL wine.GEN  
 “drank five big bottles of wine” (Babby 1988: 289)

Thus, as Babby describes matters, D itself is uniformly inflected for DP-external case. When D carries an externally determined oblique case feature, the NP head and modifiers of it must check this case. But when D carries an external, structural Case feature, D's own inherent case (genitive) wins out.

Compare now a famous phenomenon first observed by Bopp (1848) in Georgian examples like (86a). The noun *mter-ta-sa*, “of the enemies”, shows BOTH the external case marking of the head itself (DAT) AND the internal case marking (OBL.PL) relevant to its relation to the head (*çqoba* “attack”). Other examples from Bopp are given in (86b), (86c); (86d) is a parallel example from Old Georgian due to Bork (1905);

- (86) a. *çqoba-sa mter-ta-sa*  
 attack-DAT enemy-OBL.PL-DAT  
 “at the attack of the enemies”  
 b. *gwam-isa krist-es-isa*  
 body-GEN Christ-GEN-GEN  
 “of the body of Christ”  
 c. *qeli-ta mocikul-ta-ta*  
 hand-OBLPL apostle-OBL.PL-OBL.PL  
 “through the hands of the apostles”  
 d. *pir-isa-gan uymrto-ta-sa* (Bork 1905)  
 face-GEN-from infidel-OBL.PL-DAT  
 “from the face of the infidels”

This ‘double case’ phenomenon, later termed *Suffixaufnahme* by Finck (1910), occurs primarily in the situation where the Russian homogeneous agreement pattern appears according to Plank (1995). That is, in situations of oblique external case marking – dative, locative, instrumental, genitive – we get DP-internal case effects as well.

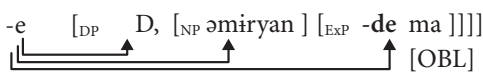
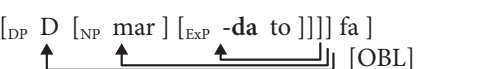
Larson (2018) proposes that Zazaki ‘doubled’ or ‘strengthened’ *Ezafe* is in fact an instance of the *Suffixaufnahme* or double case phenomenon.<sup>28</sup> Recall that doubled *Ezafe* occurs in two circumstances. The first is when one *Ezafe* construction is embedded inside another, as in (87):

- (87) a. [HEAD=EZ [HEAD=**de**/**da** MOD]]  
 b. *kutik=e* [*əmiryan=**de** ma*]  
 dog-EZ neighbor(OBL)-SEZ us  
 “our neighbor’s dog”  
 c. *ma=y* [*mar=**da** ay*]  
 mom=EZ mom(OBL)-SEZ her  
 “her mother’s mother”  
 d. *aqil-e* [*mar’dim=**de** pil-I*]  
 wisdom=EZ people=SEZ older-PL  
 “the wisdom of older people”

The second is when an *Ezafe* construction is governed by an oblique preposition, as in (82) (repeated below as (88)):

- (88) a. [HEAD=**de**/**da** MOD] P  
 b. [*embaz=**de** xwi*] *-re*  
 friend=SEZ own -to  
 “to his friend”  
 c. [*mar=**da** to*] *fa*  
 mom(OBL)=SEZ you(OBL) from  
 “from your mother”

Suppose that *Ezafe* has the status of an oblique case-marker, as postulated by the case-marking analysis. Then in both instances we are seeing *Ezafe* under an oblique case-marker – in brief, oblique under oblique. This is precisely the situation where the *Suffixaufnahme* phenomenon arises: morphology that reflects the oblique external case of the DP and the internal case of DP taken together. Larson (2018) suggests specifically that Zazaki double *Ezafe* forms *-de* and *-da* are in fact portmanteaus of the *Ezafe* element and a general oblique case coming from without, as shown in (89):

- (89) a. [<sub>EXP</sub> -e [<sub>DP</sub> D, [<sub>NP</sub> əmiryan ] [<sub>EXP</sub> -**de** ma ]]]]  
  
 b. [<sub>PP</sub> [<sub>DP</sub> D [<sub>NP</sub> mar ] [<sub>EXP</sub> -**da** to ]]]] *fa* ]  


28. This conclusion is independently reached by Plank (p.c.) in unpublished research notes.



Thus the case-marking analysis allow us to draw the otherwise idiosyncratic behavior of Zazaki ‘doubled’ *Ezafe* into a much broader picture.

### 3.4.3.2 Caspian ‘Reverse Ezafe’

As noted by Larson (2009), in the Caspian languages Mazanderani, Gilaki and Talyshi, nominals show a pattern that is nearly the mirror inverse of that found in iPersian. Thus attributive nouns, attributive adjectives, possessives, and a whole range of noun complements occur preminally, and link to N via an invariant ‘Reverse *Ezafe*’ particle (REZ), which again cliticizes to the preceding element (90):<sup>29</sup>

- (90) a. NP/AP/PP =REZ N  
 b. NP =REZ A  
 c. NP =REZ P

These patterns are illustrated in (91)–(93) for Gilaki and (94)–(95) from the Sari dialect of Mazanderani.<sup>30</sup>

Gilaki

#### (91) Modifiers & complements of Ns

- |    |                               |                 |
|----|-------------------------------|-----------------|
| a. | <i>bay=ə gul-an</i>           | NP=REZ N        |
|    | garden=REZ flower-PL          |                 |
|    | “garden flowers”              |                 |
| b. | <i>John=ə xowne</i>           | NP=REZ N        |
|    | John=REZ house                |                 |
|    | “John’s house”                |                 |
| c. | <i>ab=ə xurdan</i>            | NP=REZ N        |
|    | water=REZ eat                 |                 |
|    | “drinking of water”           |                 |
| d. | <i>surx=ə gul</i>             | AP=REZ N        |
|    | red=REZ flower                |                 |
|    | “red flower”                  |                 |
| e. | <i>xayli kushtay(=ə) utaq</i> | AP=REZ N        |
|    | very small(=REZ) room         |                 |
|    | “very small room”             |                 |
| f. | <i>xujir=ə sabz=ə kitaab</i>  | AP=REZ AP=REZ N |
|    | good=REZ green=REZ book       |                 |
|    | “good green book”             |                 |

29. The term ‘Reverse Ezafe’ appears to have been coined by Don Stilo.

30. We thank Bardyaa Hessam (p.c.) for the Gilaki data in (91)–(93) and Zia Khoshsirat (p.c.) for discussion of this and other Gilaki data. The Mazanderani examples in (94) and (95) are taken from Yoshie (1996).

- g. *daryaa(=ə) kinaar=ə xowne* [NP=REZ P]=REZ N  
 sea(=REZ) next=REZ house  
 “house beside the sea”
- (92) **Complements of As**
- a. *Hæsæn=ə aashiq* NP=REZ A  
 Hasan=REZ in love  
 “in love with Hasan”
- b. *zak=ə negarown* NP=REZ A  
 child=REZ worried  
 “worried about the child”
- c. *Gudut=ə muntazir* NP=REZ A  
 Godot=REZ waiting  
 “waiting for Godot”
- (93) **Complements of Ps**
- a. *divaar=ə sar* NP=REZ P  
 wall=REZ top  
 “up the wall”
- b. *otaq=ə vəsæt* NP=REZ P  
 center=REZ room  
 “in the middle of the room”
- c. *istaxr=ə dowri* NP=REZ P  
 pool=REZ around  
 “around the pool”
- d. *daryaa(=ə) kinaar=ə xowne* [NP=REZ P]=REZ N  
 sea(=REZ) next=REZ house  
 “house beside the sea”

## Mazanderani (Sari)

- (94) **Modifiers & complements of Ns**
- a. *dār=ə sar* NP=REZ N  
 tree=REZ top  
 “top of the tree”
- b. *asb=ə kale* NP=REZ N  
 horse=REZ head  
 “horse’s head”
- c. *farhād=ə xāxer=ə hemsāye* NP=REZ NP=REZ N  
 Farhad=REZ sister=REZ neighbor  
 “neighbor of Farhad’s sister”
- d. *me berār=ə rafeq=ə ketāb* NP=REZ NP=REZ N  
 1SG brother=REZ friend=REZ book  
 “book of my brother’s friend”

- e. *gat=ə sere* AP=REZ N  
big=REZ house  
“big house”
- f. *belend=ə ku* AP=REZ N  
high=REZ mountain  
“tall mountain”
- g. *kučik=ə ‘otāq* AP=REZ N  
small=REZ room  
“small room”
- h. *lāqer=ə sefid-ru=ə zenā* AP=REZ AP=REZ N  
thin=REZ pale-face=REZ woman  
“thin, pale-faced woman”
- (95) **Complements of Ps**
- a. *dār=ə ben* NP=REZ P  
tree=REZ under  
“under a tree”
- b. *me ‘otāq=ə dele* NP=REZ P  
1SG room=-REZ in  
“in my room”
- c. *me ‘berā=ə dembāl* NP=REZ P  
1SG brother=REZ after  
“after my brother”

An interesting departure from symmetry vis-à-vis iPersian occurs with relative clauses. Caspian reduced, nonfinite relatives (RRCs) appear prenominally bearing -REZ like other modifiers (96a)–(96a’). By contrast, Caspian finite relatives (FRCs) occur POST nominally and are introduced by complementizer (*ke*), just like those in iPersian, and show no *Ezafe*-type element (96b)–(96b’). The Mazanderani examples in (97) and (98) illustrate this difference.<sup>31,32</sup>

31. We thank Mohsen Mahdavi Mazdeh (p.c.) for the Mazanderani data in (97)–(98) and for very helpful discussion.

32. Larson (2009) gives the Gilaki (i) as a potential example of a prenominal reduced relative clause with REZ. The analysis is not straightforward, however, since Gilaki past participles end in a final -ə that is homophonous with REZ.

(i) *‘i suyis=e ji vagars=ə juvon*  
this SW=REZ from back-turn=REZ?/PP? youth  
“this young person returned from Switzerland”

The situation appears clearer in Mazanderani. M. Mazdeh (p.c.) notes that in Amol, Babol, and Nur, past participles do not generally end in a vowel (ia)–(d), hence in prenominal environments the final -ə appearing on the participle can be identified as -REZ, not PP (iia)–(c).

- (96) **CASPIAN** *iPERSIAN*  
 a. RRC =REZ N a'. N =EZ RRC  
 b. N FRC b'. N FRC
- (97) a. [tæf=ə sær bæpət ]=ə pəla<sup>33</sup> RRC=REZ N  
 fire=REZ on cooked.PPRT =REZ rice  
 “the rice cooked over a fire”  
 b. *unta pəla [kə mən tæf=ə sær bæpət-əmə]* N FRC  
 DEM.DIST rice REL 1SG fire=REZ on cooked-1SG  
 “the rice that I cooked over a fire”
- (98) a. [u=ə d̂ʒa bæfurd ]=ə peræn RRC=REZ N  
 Water=REZ with washed.PPRT -REZ shirt  
 “the shirt washed with water”  
 b. *unta peræn [kə tə u=ə d̂ʒa bæfurd-i]* N FRC  
 DEM.DIST shirt REL 2SG water=REZ with washed-2SG  
 “the shirt that you washed with water”

- 
- (i) a. *vənə ling bəfkəs biə.*  
 her/his/its leg **broken** was  
 “Her/his/its leg was broken”  
 b. *inta mast hi bæχərd niə.*  
 this yogurt stir **eaten** is.not  
 “This yogurt is not stirred.”  
 c. *pəla bæpət biə.*  
 rice **cooked** was  
 “the rice was cooked”  
 d. *peræn bæfurd biə.*  
 shirt **washed** was  
 “the shirt was washed.”
- (ii) a. *bəfkəss=ə ling*  
**Broken=REZ** leg  
 “broken leg”  
 b. *hi bæχərd=ə mast*  
 stir **eaten=REZ** yogurt  
 “stirred yogurt”  
 c. *bənə bæχərd=ə adəm*  
 ground **hit=REZ** person  
 “person who has fallen down”

Mazdeh also observes that the stress patterns of the Mazanderani participles in (i) and (ii) are distinctive. Whereas Mazandarani verbs always stress the preverb (*bə* or *bæ*) when there is one, stress in the participle is always on the last syllable (*kəs* in *bəfkəs* and *χərd* in *bæχərd*). He notes that this gives us further confirmation that these forms are nouns, not verbs.

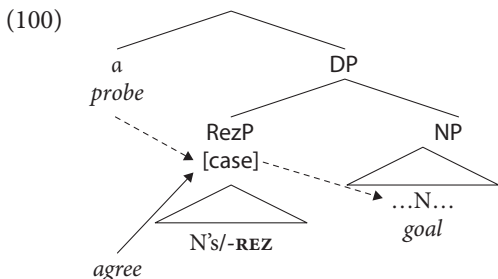
33. Note that (97a) and (98a) both involve participles that do not end in a vowel; cf. fn.28 (ic)–(id). Hence -ə in (97a) and (98a) is unambiguously -REZ.

Larson (2009) proposes that Caspian Reverse *Ezafe* also reflects case. Specifically, Larson suggests that whereas *Ezafe* represents a generalization of the case relations found with English *of* and iPersian *az*, Reverse *Ezafe* represents a generalization of case relations found with English 's. Corbett 1987 observes that various Slavonic languages contain suffixes for creating 'possessive adjectives' from nouns. For example, in addition to familiar post nominal genitives like (99a), Upper Sorbian (spoken in Lusatia, eastern Germany, has possessive adjectives, formed by suffixing *-in/-yn* to feminine nouns and *-ow* to masculine nouns (99b).

- (99) a. *kniha Jan-a* Upper Sorbian (Corbett 1987)  
 book Jan-GENSG  
 "a/the book of Jan's"  
 b. *Jan-ow-a* *knih-a*  
 Jan-POSS-NOMSGFEM book-NOMSGFEM  
 "Jan's book"

As Corbett (1987) notes the possessive *Janowa* in (99b), although derived from a masculine noun, is adjectival in behavior; thus it precedes the head and shows the same agreement as an attributive adjective – here agreeing with the nominative, feminine singular head *knih-a* "book"). Larson (2009) refers to morphemes like Upper Sorbian *-in/-yn/-ow*, which derive adjectival/concordial forms from Ns as 'concordializers'. In essence, a concordializer converts an expression requiring case by assignment – a 'nominal' – to one allowing case by agreement – an 'adjectival'.


Larson (2009) analyzes English prenominal genitive 's and Caspian Reverse *Ezafe* morphemes as a concordializers. They allow the nominal expression to which they attach to obtain case by agreement with a higher case-probe  $\alpha$  when the latter comes into agreement with the nominal head. In order to obtain such agreement, 's and REZ-marked phrases must position themselves in prenominal position, between the case probe and its goal (100):




Under these proposals, *Ezafe* and Reverse *Ezafe* pattern together as alternative general strategies for solving the same syntactic problem: how is a [+N] XP complement or modifier of a noun to satisfy its case-requirements? The *Ezafe* strategy introduces

an additional case probe into the derivation (101a), solving the problem by direct assignment or checking. The Reverse *Ezafe* strategy introduces a concordializer into the derivation (101b), solving the problem by agreement.

- (101) a. **-EZ Checks Case on [+N] XP**  

$$N \quad [{}_{\text{EzP}} = \text{EZ} \quad \text{XP}]$$

*assignment/checking*
- b. **-REZ Concordializes [+N] XP**  

$$[{}_{\text{RezP}} \text{XP} = \text{REZ}] \quad N$$

*agreement*

Constructions of the first sort would be complement-like; constructions of the second sort would be fundamentally attributive in nature.

#### 4. Concluding remarks

The case-marking analysis offers an approach to *Ezafe* and Reverse *Ezafe* distribution that is more adequate in empirical coverage and richer in theoretical predictions than competitors. It also carries interesting typological implications about the kind of language that manifests *Ezafe*/Reverse *Ezafe* phenomena and why they do so. It seems to us that the crucial parameters at work in languages of the relevant sort must concern the case properties of adjectives and prepositions and how they align with nouns (Karimi & Brame 1986, 2012). Under usual views, nouns are referential, denote properties (e.g., beauty, truth), occur as arguments and are assigned (or valued for) case. By contrast, adjectives are non-referential, denote predicates (e.g., beautiful, true), occur as attributive modifiers and agree for case. Prepositions are non-referential, typically denote relations (e.g., in, before), occur as attributive modifiers and are neither assigned case nor agree for case.

These N-A differences are evident in English in contrasts like (102)–(104). (102) shows that Ns but not As can occur in argument position. (103) shows that Ns are not freely substitutable for their corresponding As in attributive position. Finally (104) shows that As are not freely substitutable for Ns as P-objects:

- (102) a. *We discussed truth/beauty.* Argument position  
 b. \**We discussed true/beautiful.*
- (103) a. *A very long/\*great length road* Attributive position  
 b. *A very important /\*great importance article*  
 c. *A very thick/\*great thickness book*

- (104) a. *A road of great length/\*very long* Object of P  
 b. *An article of great importance/\*very important*  
 c. *A book of great thickness /\*very thick*

iPersian seems to exhibit the same distributional facts. Ns but not As are permitted in argument positions (105) (cf. (102)); Ns are not freely substitutable for As in attributive constructions (106) (cf. (103)), and As are not freely substitutable for Ns as objects of Ps (107) (cf. (104)).

- (105) a. *Ma raje be hæghighæt/zibayi bæhs=kærd-im.*  
 we about to truth/beauty discussion did-1PL  
 “We discussed truth/beauty.” Argument position
- b. *\*Ma raje be hæghighi/ziba bæhs=kærd-im.*  
 we about to true/beautiful discussion did-1PL  
 “We discussed true/beautiful.”
- (106) a. *Ye jade=ye besyar tulani/\*[tul=e ziad]* Attributive position  
 INDF road=EZ very long /\*length=EZ great  
 “a very long road”
- b. *Ye mæqale=ye [besyar mohem] /\*[æhæmiæt=e ziyad]*  
 INDF article=EZ very important /\*importance=EZ great  
 “a very important article”
- c. *Ye ketab=e [besyar zækhim] /\*[zekhamæt=e ziad]*  
 INDF book=EZ very thick /\*thickness=EZ great  
 “a very thick book”
- (107) a. *Ye jade ba [tul=e ziad]/\*[besyar tulani]* Object of P  
 INDF road with length=EZ great/\*very long  
 “a road of great length”
- b. *Ye mæqale(=ye) ba [æhæmiæt=e ziyad]/\*[besyar mohem]*  
 INDF article(=EZ) with importance=EZ great /\*very important  
 “an article of great importance”
- c. *Ye ketab ba [zekhamæt=e ziad]/\*[besyar zækhim]*  
 INDF book with thickness-EZ great/\*very thick  
 “a book of much thickness”

What differences there are would therefore not seem to be ‘deep’ ones, wherein adjectives in iPersian are actually nouns (contra Karimi & Brame 1986, 2012). Rather the difference would seem to be more superficial, regarding how case features are realized with [+N] items in the two languages. In English (and many other languages), As are concordial for case; i.e., case is an uninterpretable/unvalued feature. By contrast, in *Ezafé* and Reverse *Ezafé* languages, adjectives appear to behave featurally like nouns; i.e., case is a valued feature on both As and Ns. This result in turn

suggests that concordiality/agreement is not a ‘deep’ syntactic property of adjectives and what is normally taken to be the usual situation, with adjectives agreeing with their nouns, is in fact a derived one, involving more structure than is typically assumed.<sup>34</sup> We must leave these intriguing speculations for future exploration.<sup>35</sup>

## References

- Anderson, Stephen. 2013. The marker of the English ‘Group Genitive’ is a special clitic, not an inflection. In Kersti Börjars, David Denison & Alan Scott (eds.), *Morphosyntactic categories and the expression of possession*, 193–217. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/la.199.o8and>
- Aghaei, Behrad. 2006. The Syntax of ke-clause and clausal extraposition in Modern Persian. Austin, TX: University of Texas at Austin dissertation.
- Babby, Leonard. 1987. Case, prequantifiers, and discontinuous agreement in Russian. *Natural Language and Linguistic Theory* 5. 91–138. <https://doi.org/10.1007/BF00161869>
- Babby, Leonard. 1988. Noun phrase internal case agreement in Russian. In Michael Barlow & Charles A. Ferguson (eds.), *Agreement in natural language*, 287–304. Stanford, CA: CSLI Stanford.
- Barwise, Jon & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4. 159–219. <https://doi.org/10.1007/BF00350139>
- Bopp, Franz. 1848. Über das Georgische in sprachverwandtschaftlicher beziehung. *Abhandlungen der Königlichten Akademie der Wissenschaften zu Berlin, philosophisch- historische Klasse*, 1846. 259–339.
- Bork, Ferdinand. 1905. Kaukasisches. *Orientalistische Litteratur-Zeitung* 8. 184–187.
- Bowers, John. 1993. The syntax of predication. *Linguistic Inquiry* 24. 591–656.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 2010. *The syntax of adjectives: A comparative study*. Cambridge, MA: MIT Press.  
<https://doi.org/10.7551/mitpress/9780262014168.001.0001>
- Citko, Barbara. 2004. On headed, headless, and light-headed relatives. *Natural Language and Linguistic Theory* 22. 95–126. <https://doi.org/10.1023/B:NALA.0000005564.33961.e0>
- Corbett, Greville G. 1987. The morphology/syntax interface: Evidence from possessive adjectives in Slavonic. *Language* 63. 299–344. <https://doi.org/10.2307/415658>
- Den Dikken, Marcel. 2006. *Relators and linkers: A study of predication, predicate inversion and copulas*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/5873.001.0001>

---

34. Larson (2018) argues for a similar conclusion regarding adjectives in Mandarin.

35. Parallel issues arise with the category P. It is well known that PPs, particularly locatives, can behave referentially and hence nominally in natural language (*Under the bed seems to be a good place to hide.*). Furthermore many locative Ps (*behind, beneath, in-the-midst-of*, etc.) are recognized to possess a nominal core (*hind, neath, midst*, etc.). Hence alignment of P with N poses similar issues.



- Den Dikken, Marcel & Pornsiri Singhapreecha. 2004. Complex noun phrases and linkers. *Syntax* 7. 1–54. <https://doi.org/10.1111/j.1368-0005.2004.00064.x>
- Downing, Bruce. 1978. Some universals of relative clause structure. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language, volume 4: Syntax*, 375–418. Stanford, CA: Stanford University Press.
- Emonds, Joseph. 1976. *A transformational approach to English syntax: Root, structure-preserving, and local transformations*. New York: Academic Press.
- Finck, Franz Nikolas. 1910. *Die Haupttypen des Sprachbaus*. Leipzig: Teubner.  
<https://doi.org/10.2307/199814>
- Franco, Ludovico, M. Rita Manzini & Leonardo Savoia. 2015. Linkers and agreement. *The Linguistic Review* 32. 277–332.
- Givón, Talmay. 1990. *Syntax: A functional-typological introduction*, vol. II. Amsterdam: John Benjamins.
- Göksel, Asli & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London: Routledge.
- Haig, Geoffrey. 2011. Linker, relativizer, nominalizer, tense-particle: On the Ezafe in West Iranian. In Foong Ha Yap, Karen Grunow-Härsta & Janick Wrona (eds.), *Nominalization in Asian languages: Diachronic and typological perspectives*, 363–390. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/tsl.96.13hai>
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden, MA: Wiley.
- Jackendoff, Ray. 1973. The base rules for prepositional phrases. In Stephen R. Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 345–356. New York: Holt, Rinehart and Winston.
- Jackendoff, Ray. 1977. *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.
- Kahnemuyipour, Arsalan. 2014. Revisiting the Persian Ezafe construction: A roll-up movement analysis. *Lingua* 150. 1–24. <https://doi.org/10.1016/j.lingua.2014.07.012>
- Karimi, Simin & Michael Brame. 1986. A generalization concerning the Ezafe construction in Persian. *West Coast Conference in Linguistics (WECOL 86)*, Vancouver, Canada.
- Karimi, Simin & Michael Brame. 2012. A generalization concerning the Ezafe construction in Persian. *Linguistic Analysis* 38. 111–144.
- Karimi, Yadgar. 2007. Kurdish Ezafe construction: Implications for DP structure. *Lingua* 117. 2159–2177. <https://doi.org/10.1016/j.lingua.2007.02.010>
- Klima, Edward. 1965. *Studies in diachronic syntax*. Cambridge, MA: Harvard University dissertation.
- Krause, Cornelia. 2001. *On reduced relatives with genitive subjects*. Cambridge, MA: MIT dissertation.
- Lapointe, Steven. 1990. EDGE features in GPSG. *Proceedings of the Chicago Linguistic Society* 26. 221–235. Chicago, IL: University of Chicago.
- Lapointe, Steven. 1992. Life on the EDGE: Arguments in favor of an autolexical account of edge inflections. *Proceedings of the Chicago Linguistic Society* 28. 318–332. Chicago, IL: University of Chicago.
- Larson, Richard and Gabriel Segal. 1995. *Knowledge of meaning*. Cambridge, MA: MIT Press.  
<https://doi.org/10.7551/mitpress/4076.001.0001>
- Larson, Richard & Hiroko Yamakido. 2008. Ezafe and the deep position of nominal modifiers. In Louise McNally & Christopher Kennedy (eds.), *Adjectives and adverbs: Syntax, semantics and discourse*, 43–70. Oxford: Oxford University Press.
- Larson, Richard. 2009. Chinese as a reverse Ezafe language. *Yuyanxue Luncong [Essays on Linguistics]* 39. 30–85.

- Larson, Richard. 2018. Zazaki 'double *Ezafe*' as double case-marking. Manuscript, Stony Brook University.
- Larson, Richard. 2018. AP-地 adverbs in Mandarin. *Studies in Chinese Linguistics* 39. 1–28. <https://doi.org/10.2478/scl-2018-0001>
- Larson, Richard & Vida Samiiian. 2018. *Ezafe*, PP and the nature of nominalization. Paper presented at the 54th Meeting of the Chicago Linguistic Society. Chicago, IL: University of Chicago.
- Larson, Richard, Elias Abdollahnejad, Mohammad Ghaderibaneh, Songül Gündoğdu, Ali Salehi & Vida Samiiian. 2019. Turkish influence in Iranian relative clause syntax? Paper presented at Tu+4, NYU, February 17, 2019.
- Lavine, James. 2003. Resumption in Slavic: Phases, cyclicity and case. In Mila Tasseva-Kurkchieva, Steven Franks & Frank Y. Gladney (eds.), *Formal Approaches to Slavic Linguistics* 13. 195–206. Ann Arbor, MI: Michigan Slavic Publications.
- Lehmann, Christian. 1986. On the typology of relative clauses. *Linguistics* 24. 663–680. <https://doi.org/10.1515/ling.1986.24.4.663>
- Lehmann, Christian. 1988. Toward a typology of clause linkage. In John Haiman & Sandra A. Thompson (eds.), *Clause combining in grammar and discourse*, 181–226. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.18.09leh>
- Manzini, Rita. 2010. The structure and interpretation of (Romance) complementizers. In Phoevos Panagiotidi (ed.), *The complementiser phase: Subjects and operators*, 167–199. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199584352.003.0008>
- McCloskey, James. 2002. Resumption, successive cyclicity, and the locality of operations. In Samuel Daniel Epstein & T. Daniel Seely. (eds.), *Derivation and explanation in the minimalist program*, 184–226. Oxford: Blackwell. <https://doi.org/10.1002/9780470755662.ch9>
- Merchant, Jason. 2004. Resumptivity and non-movement. *Studies in Greek Linguistics* 24. 471–481. Thessaloniki: Aristotle University of Thessaloniki.
- Miller, Philip. 1991. Clitics and constituents in phrase structure grammar. Rijksuniversiteit te Utrecht dissertation.
- Nevis, Joel. 1986. Finnish particle clitics and general clitic theory. *Working Papers in Linguistics* 33. Columbus, OH: Department of Linguistics, The Ohio State University.
- Palmer, Adrian. 1971. The *Ezafe* construction in Modern Standard Persian. Ann Arbor, MI: University of Michigan dissertation.
- Pesetsky, David & Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In Simin Kairimi, Vida Samiiian & Wendy K. Wilkins (eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation*, 262–294. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.101.14pes>
- Pikkert, Peter. 2010. *A basic course in Kurmanji*. Ontario: Alev Books.
- Plank, Frans. 1995. (Re-)Introducing Suffixaufnahme. In Frans Plank (ed.), *Double case*, 3–110. New York: Oxford University Press.
- van Riemsdijk, Henk. 1990. Functional prepositions. In Harm Pinkster & Inge Genée (eds.), *Unity in diversity*, 229–241. Dordrecht: Foris. <https://doi.org/10.1515/9783110847420.229>
- Samiiian, Vida. 1983. Structure of phrasal categories in Persian: An X-bar analysis. Los Angeles, CA: UCLA dissertation.
- Samiiian, Vida. 1994. The *Ezafe* construction: Some implications for the theory of X-bar syntax. In Mehdi Marashi (ed.), *Persian Studies in North America*, 17–41. Bethesda, MD: Iranbooks.
- Samvelian, Pollet. 2007. A (phrasal) affix analysis of the Persian *Ezafe*. *Journal of Linguistics* 43. 605–645. <https://doi.org/10.1017/S0022226707004781>

- 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. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Schmidtke-Bode, Karsten. 2014. Complement clauses and complementation systems: A cross-linguistic study of grammatical organization. Friedrich-Schiller-Universität Jena dissertation.
- Svenonius, Peter. 2003. Limits on P: Filling in holes vs. falling in holes. *Nordlyd, Tromsø Working Papers on Language and Linguistics. Proceedings of the 19th Scandinavian Conference of Linguistics*, 31, 431–445.
- Todd, Terry Lynn. 1985. A grammar of Dimili (also known as Zaza). Ann Arbor, MI: University of Michigan unpublished dissertation.
- Yoshie, Satoko. 1996. *Sari dialect (of Mazandarani language)*. Tokyo, Japan: Institute for the Study of Languages and Cultures of Asia and Africa.
- Zwicky, Arnold. 1987. Suppressing the Zs. *Journal of Linguistics* 23. 133–148.  
<https://doi.org/10.1017/S0022226700011063>

# Quantitative meter in Persian folk songs and pop lyrics

Mohsen Mahdavi Mazdeh

University of Arizona

This chapter argues that, contrary to what most recent scholarly works assume, the metrical system used in Persian folk songs and pop lyrics is quantitative and follows the same general principles as Classical Persian metrics. I propose that the apparent differences between the two systems originate primarily from the availability of a process of optional vowel shortening in the scansion of lines that are composed in colloquial Persian. In fact, it is mainly the phonological differences between the colloquial and formal registers of Persian, rather than purely metrical differences, that result in the split observed between these two poetic traditions. In addition to optional vowel shortening, I describe several minor deviations in these songs from the requirements of Classical Persian metrics, showing that these deviations are also systematic and that studying them can help gain a deeper understanding of Persian metrics. Finally, I present corpus data to support these proposals.

**Keywords:** Persian meters, quantitative meter, folk songs, pop lyrics, vowel length, colloquial Persian, spoken Persian, vowel shortening

## 1. Introduction

Classical Persian poetry is known to have a quantitative metrical system based on syllable weight (Hayes 1979; Deo & Kiparsky 2011; Najafi 2015). However, such a consensus does not exist for colloquial Persian songs. Currently the dominant view on the metrical structure of these songs is that introduced by Tabibzadeh (2003) for folk songs, and expanded by Azarmakan & Nejati Jazeh (2014) and Tabibzadeh & Mirtalae (2015) to cover children's songs and pop lyrics, respectively. According to this view, the meters of the poems of these songs rely on stress and are not quantitative.

I argue that the metrical structure of Persian folk songs and pop lyrics is quantitative and in fact quite similar to that of Classical Persian poetry. In particular, the

possibility of a short reading for the traditionally long Persian vowels in colloquial Persian is at the core of the difference between the two metrical systems. I propose that this process of vowel shortening may reflect the phonological properties of colloquial Persian, the variety of Persian that is used in folk songs and pop lyrics.

The remainder of this chapter is structured as follows. In § 2, I present a brief account of the metrical system of Classical Persian poetry as it is analyzed in the literature today. In § 3, I lay out the details of my proposal on Persian folk songs and pop lyrics, based on the pivotal claim that they follow the same general principles as Classical Persian poetry. In § 4, I review the main alternative theory that has been presented in the literature for analyzing the metrical system of Persian folk songs and pop lyrics, arguing that it suffers from both theoretical and empirical problems. In § 5, I present corpus data to support my proposal, followed by a conclusion section in which I review the current proposal and its theoretical implications.

## 2. Meters in Classical Persian poetry

In Classical Persian poetry, meters are based on syllable weight, which is in turn based on the number of moras in a syllable. Following Hayes (1979), I view the metrical system of Classical Persian as consisting of two major components. The first component is composed of the correspondence criteria that determine what sequence(s) of light and heavy syllables any given Persian phrase can correspond to. The second component, to which I refer as the metrical assessment process, determines which sequences of light (L) and heavy (H) positions are metrical. In the example given in (1), for example, the correspondence criteria determine that each of the lines can be mapped to the syllable sequence LLHLLHLLH (more on how this is determined in § 2.1). The metrical assessment process then comes into play, evaluating the sequence LLHLLHLLH (commonly fragmented as LLHH-LLHH-LLH, highlighting the repeating pattern of the syllables) as a metrically valid sequence. The combination of the two components, therefore, determines that the lines given in (1) are metrical, and follow the same meter.

- (1) LL    HHL    L HH    L LH  
*mæn-e bi:tʃa:re    je gærdæn be kæmænd*  
 I-e    hopeless    e neck    to  lasso  
 “I, the hopeless captive”  
 L    LH    H L LHH    LLH  
*tʃe konæm gær be reka:b-æf    nærævæm*  
 what I.do    if    to ride-his/her I.do.not.go  
 “What can I do but to follow his/her lead?”

(Saadi, Ghazal 427, line 6)

Systematically describing and finding the principles behind the metrical assessment process is the subject of ongoing research (e.g., Deo & Kiparsky 2011; Najafi 2015; Mahdavi Mazdeh 2019). The correspondence criteria, however, are fairly straightforward and there is consensus over them in the literature (see Hayes 1979; Shamisa 2004; Thiesen 1982). In the following subsection, I present a brief overview of the most important aspects of the correspondence criteria in Classical Persian poetry, before moving to a discussion of meter in folk songs and pop lyrics in § 3.

## 2.1 Correspondence criteria in Classical Persian poetry

Whether a line can be mapped to a particular metrical pattern or not is primarily determined by the weights of its syllables, i.e. how many moras they have. The mora count of a syllable is determined as follows. Coda consonants and short vowels (/æ/, /e/, and /o/) are moraic, and long vowels (/a/, /u/, and /i/) are bimoraic. From different combinations of vowel lengths and coda counts, three levels of syllable weight emerge, as shown in (2).

### (2) Syllable weight in Classical Persian poetry

light (L):	CV	e.g. /to/
heavy (H):	CVC, CVV	e.g. /tor/, /taː/
superheavy (S):	CVCC, CVVC, CVVCC	e.g. /tord/, /taːb/, /taːft/

Notice that CVVCC syllables are analyzed as superheavy (trimoraic) even though a superficial mora count on its structure suggests four moras. This seemingly unexpected but well-known phenomenon in Persian metrics is something I assume in this chapter (for discussion, see Hayes 1979).

Superheavy syllables are always interchangeable with a sequence composed of a heavy syllable followed by a light syllable (HL) in line-medial positions, and with a heavy syllable (H) in line-final positions. As a result, metrical patterns are identified only with L and H symbols, and the correspondence criteria map the superheavy syllables in verses to either HL or H depending on their positions.

With these correspondence criteria, linguistic phrases can be mapped to syllable sequences. In (1), for example, the final syllable of the first line (/mænd/) is a superheavy syllable, but it is mapped to an H since it is in line-final position, in accordance with the criteria mentioned above.

An exception to the general rules is that in the presence of /n/ as the first coda consonant, the sequence CVVn is treated like CVn (the vowel is in fact pronounced shorter). Another occasion in which a long vowel has to be counted (and pronounced) as short is the case of prevocalic /i/. Short vowels, on the other hand, are ALLOWED (but not required) to be parsed as long in word-final position. Finally,

a meter-specific poetic license on the meter (LLHH-LLHH-LLH) allows the first syllable to appear as heavy.

The points introduced above are demonstrated in Example (3). The meter is the same as the previous example (LLHH-LLHH-LLH).

- (3) H L HH LLH HHH  
*in ke χa:k-e: siæh-æf ba:li:n-æst*  
 this that soil-*e* black-his/her bed-is  
 “She, who now has the dark earth as her bed”  
 HLH HL LH HHH  
*æχtær-e: tʃærχ-e ædæb pærvi:n-æst*  
 star-of sky-of literature parvin-is  
 “Is Parvin, a star in the sky of literature” (Etesami, tombstone poem, line 1)

In both verses, the first syllable is heavy but is licensed to correspond to an L position as discussed above, and the last syllable is superheavy but corresponds to a heavy position (by convention, only L and H syllables are used in glossing). Moreover, the prevocalic /i/ in the word /siæhæf/ and the /i/ before coda /n/ in the word /in/ are both parsed as short. In order for the lines to conform to the intended meter, the use of the poetic license allowing word-final short vowels to be parsed as long needs to be assumed for /χa:ke:/ and /æχtære:/, but not for /ke/.

Now that the most important aspects of the correspondence criteria for mapping actual lines to abstract syllable sequences in Classical Persian poetry have been introduced, we can turn to folk songs and pop lyrics. In the metrical scansion presented in the remainder of this chapter, no aspects of the correspondence criteria other than the ones mentioned above are referred to.

### 3. Meters in Persian folk songs and pop lyrics

In the last few decades, the dominant view in the literature on Persian folk songs and pop lyrics has been that these poems follow an entirely non-quantitative metrical system based merely on stress and syllable count. This particular view, which is inspired by older works such as Khanlari (1948) and Adib Toosi (1953), is introduced by Tabibzadeh (2003) initially for Persian folk songs, and expanded later to cover Persian pop lyrics (Tabibzadeh & Mirtalaei 2015) as well as songs in Gilaki, another Iranian language (Tabibzadeh 2010), and Persian children’s songs (Azarmakan & Nejati Jazeh 2014). This view is discussed in more detail in § 4.

I argue for an alternative proposal, which considers the meter in these forms of poetry to be purely quantitative, and based on the same metrical system as Classical Persian poetry. This idea has been proposed before by Vahidian Kamyar (1979), but he does not provide a systematic explanation for why folk songs and pop lyrics

seem to violate the metrical rules of Classical Persian poetry. In the absence of such an explanation, the competing theory has gained ground in recent decades.

### 3.1 Optional vowel shortening

I propose that the distinction between the two components of a metrical system, i.e. the correspondence criteria and the metrical assessment process, is the key factor that needs to be taken into consideration for settling the issue. For instance, consider the lines in (4), taken from the lyrics of a well-known contemporary pop song.

- (4) HL        H    HL    LH    H    HLH  
*da:re    æz   æbr-e   sija:   χun   mitʃeke:*  
 PROG.3SG from cloud-*e* black blood drips  
 “Black clouds are dripping with blood”  
 HLH        H    H    L    HH    HLH  
*āzomʔeha:   χun   āza   je   barun   mitʃeke:*  
 Fridays        blood place of rain    drips  
 “On Fridays, it rains blood instead of water” (Ghanbari 1998, *Jom'eh*, 5–6)

Using the criteria normally used for parsing verses of Classical Persian poetry, the first line follows the pattern HLHH-LLHH-HLH (the dashes are only for ease of reading) and the second line follows the pattern HLHH-HHLH-HLH. While the patterns look similar to each other and identical to known metrical patterns (as Vahidian Kamyar has noticed in similar cases), they are not identical and do not match any known valid meters of Classical Persian poetry. However, these poems can be parsed as completely metrical lines if an *optional vowel shortening* process is allowed in the correspondence criteria, which allows any long vowel to be parsed as short, if needed.

Assuming that optional vowel shortening is allowed and using the poetic license discussed in the previous section that allows mapping heavy syllables to L positions verse-initially if the intended meter is LLHH-LLHH-LLH, we can get the desirable scansion as shown in (5). The three syllables for which optional vowel shortening is assumed are underlined.

- (5) HL        H    HL    LH    H    LLH  
*da:re    æz   æbr-e   sija:   χun   mitʃeke:*  
 PROG.3SG from cloud-*e* black blood drips  
 “Black clouds are dripping with blood”  
 HLH        H    L    L    HH    LLH  
*āzomʔeha:   χun   āza   je   barun   mitʃeke:*  
 Fridays        blood place of rain    drips  
 “On Fridays, it rains blood instead of water” (Ghanbari 1998, *Jom'eh*, 5–6)



To show the consistency of the metrical pattern, the scansion for the two preceding verses of the same poem are shown in (6). Note that in the last verse vowel shortening allows an otherwise superheavy syllable to be treated as heavy.

- (6) H L HH LL H HLLH  
*tu: je ga:b-e: χis-e in pændʒereha:*  
 in of frame-*e* wet-*e* this windows  
 “In the wet frames of these windows”  
 HL H HLL HH LLH  
*æks-i æz dʒomʔe-je gæmgin mibinæm*  
 image-INDF from Friday-*e* sad I.see  
 “I see an image of the sad Friday”  
 L LHH L LH HL LH  
*tʃe sia:h-e: be tæn-ef ræχt-e æza:*  
 how black-is to body-its clothes-of mourning  
 “How black is the mourning dress covering his body”  
 H LH HL L HH LLH  
*tu: tʃeʃa-f æbra je sængin mibinæm*  
 in eyes-its clouds *e* heavy I.see  
 “I see heavy clouds in its eyes” (Ghanbari 1998, *Jom'eh*, 1–4)

### 3.2 The origins of optional vowel shortening

To see why vowel shortening is allowed in the correspondence criteria of the metrical system under study, it may be useful to note that how syllables map to metrical positions is a language-specific issue. The three long Persian vowels, for example, are bimoraic in contexts other than poetic meter too in Classical Persian. It is possible for two languages to use similar metrical assessment processes (i.e. have similar meter inventories), as Classical Persian and Ottoman Turkish do (for a comparison of the systems see Thiesen 1982), but in any such case they almost certainly map different consonant and vowel combinations to the light and heavy positions of the metrical patterns.

With this in mind, an important difference between Classical Persian poetry on the one hand and Persian folk songs and pop lyrics on the other hand is that the latter class of poems are composed in colloquial Persian, and fully conform to the syntax and phonology of colloquial Persian. Optional vowel shortening, the main distinguishing factor of the metrical system under study, is in fact a reflection of the phonological properties of vowels in colloquial Persian. Hence, the fact that a number of researchers have questioned the presence of a phonological length distinction in contemporary Persian vowels (for a review of these studies and more

in-depth discussion see Toosarvandani 2004) may in fact be due to the fact that colloquial Persian (but not necessarily contemporary formal Persian) allows long vowels to be pronounced as short.

### 3.3 More deviations from Classical Persian metrics

In the vast majority of cases, optional vowel shortening is sufficient for describing meter in pop lyrics. In children's songs, however, the issue is further complicated by the fact that certain metrical flexibilities that are allowed only in limited environments in Classical Persian poetry are abundant in colloquial Persian poetry. Consider the example in (7), showing the scansion of the first 8 lines of a famous Persian children's song. All cases of vowel shortening are underlined.

- (7) LLLL    H    L    HL    H  
*hæsæni-e ma: je bærre daft*  
 Hasani-of us    one lamb had  
 "Our Hasani had a lamb"  
 HLL            HL H    LH  
*bærræ-f-o xejli dus midaft*  
 lamb-his-OBJ very friend had  
 "He loved his lamb very much"  
 HL    L    HH            LLH  
*bærre je f̂ag-o: topoli:*  
 lamb e fat-and plump  
 "A plump lamb"  
 HL            LHH            LLH  
*zebr-o zeræng-o: togoli:*  
 lively-and energetic-and cute  
 "Lively, energetic, and cute"  
 H    LLH            H    LLH  
*dæs kutfulu: pa: kutfulu:*  
 hand small    foot small  
 "Small hands and small feet"  
 HL    LH    HL    LH  
*pæsm-e tæm-ef kork-e holu:*  
 wool-of body-its fuzz-of peach  
 "Its woolly body like a fuzzy peach"  
 LH    LH    LH    LH  
*xodef sefid dom-ef sia:*  
 itself white tail-its black  
 "A white body and a black tail"

LL    LLH    HL    LH  
*sar-e kakol-ef ræng-e hæna:*  
 head-of forelock-its color-of henna

“The tip of its forelock, the color of henna (Ehterami 1987, 1:1–8)

The metrical patterns the lines are mapped to are shown in (8) in an easier-to-read format. The way the patterns are fragmented using spaces is intended to demonstrate the internal structure of the meter and is not part of the output of the mapping process.

- (8) LLLLH LHLH  
 HLLH LHLH  
 HLLH HLLH  
 HLLH HLLH  
 HLLH HLLH  
 HLLH HLLH  
 LHLH LHLH  
 LLLLH HLLH

The meter in each verse in (8) is commonly divided into two units (traditionally called ‘rukn’, later called ‘feet’ by modern scholars, e.g., Hayes 1979; Deo & Kiparsky 2011). Each of the two feet of the verse seems to be allowed to emerge as either LHLH or HLLH. The interchangeability of these two feet is well-attested in Classical Persian poetry (Shamisa 2004). However, unlike what we see in (8), its occurrence is relatively rare in Classical Persian poetry. Another important property of the patterns in (8) is that in two of the feet an H syllable is replaced by LL, resulting in LLLLH instead of HLLH. In the terminology of Latin and Greek metrics, the heavy position is resolved into an LL sequence. This type of replacement, again, is well-attested in Classical Persian poetry (Shamisa 2004), but has more limited usage, and is never used in vicinity of another L syllable (LLL sequences are avoided). These differences, together with the frequent use of optional vowel shortening, give an entirely different appearance to the metrical structure of these songs, and have resulted in scholars doubting their quantitative nature.

Another interesting fact regarding poems in colloquial Persian is that super-heavy syllables are typically avoided altogether in these poems. In most cases, CVVC syllables are read as H (the long vowel is pronounced short) and CVVCC are avoided as much as possible. For instance, in the past simple verb *daft* meaning “had”, the final /t/ is dropped and the verb is pronounced as *daf* with a shortened /a/, counting as an H syllable.

In folk songs (as opposed to pop lyrics), the deviation from Classical Persian metrics is usually even more noticeable. Even heavy syllables that are heavy due to the presence of a coda (and therefore cannot be mapped to an L position using optional

vowel shortening) are occasionally mapped to L positions in folk songs. These cases do indeed sound a little problematic to the native ear, and are relatively rare. However, these are probably more than occasional ‘errors’, and may reflect a systematic poetic license in this tradition. After all, the reverse case of using light syllables (CV syllables that are not word-final) in heavy positions is never allowed in folk songs.

The example in (9) shows this phenomenon. The song is a very popular Persian folk song that any native speaker of Iranian varieties of Persian is expected to have heard. In the first word of the third line, the heavy syllable /næm/ (shown with boldface H) is used in a light position.

- (9) L H LH HLLH  
*je tup daræm gelgeli-e:*  
 one ball I.have round-is  
 “I have a round ball”  
 HL LH-L HLH  
*sorχ-o sefi:d-o a:bi-e:*  
 red-and white-and blue-is  
 “It is red, blue, and white”  
LLH LH LH LH  
*mizænæm zæmin hæva: mire:*  
 I.hit earth air goes  
 “When I hit it on the ground, it bounces back into the air”  
 LLLL H LH LH  
*nemiduni ta: kodʒa: mire:*  
 you.do.not.know till where goes  
 “You can’t imagine how far it goes”  
 LH LH LHH  
*mæn-in tup-o: nædæftæm*  
 I-this ball-OBJ I.did.not.have  
 “I didn’t have this ball”  
 HLL H LHH  
*mæʃgam-o χub neveʃtæm*  
 my.homework-OBJ good I.wrote  
 “I did my homework well”  
 LH LH HL H  
*babam behem ejdi dad*  
 my.dad to.me present gave  
 “My dad gave me a present”  
 L HL HLH H  
*je tupp-e gelgeli: dad*  
 one ball-e round gave  
 “He gave me a round ball”

The abstract patterns these lines map to are shown in a simpler format in (10).

- (10) LHLH HLLH  
 HLLH LHLH  
 LLHLH LHLH  
 LLLLH LHLH  
 LHLH LHH  
 HLLH LHH  
 LHLH HLH  
 LHLH LHH

The meter in the first four verses is the same as the one in (8), with HLLH, LHLH, and LLLLH being used interchangeably. The only exception is LLHLH in the third verse. Even though deviations of this kind from the metrical pattern do sometimes occur in folk songs, they are indeed considered slightly malformed. Interestingly, in this particular case, some people sing an alternative version of the song replacing the word *mizænaem* (“I hit”) with *mizæni* (“you hit”) with a short final /i/, presumably to resolve the metrical malformedness.

The last four lines follow variants of a slightly different pattern (LHLH LHH), where it seems that the second foot is shortened through a process of catalexis (deletion of the last syllable). Persian only allows heavy syllables at verb final positions, so the final L after the catalexis becomes an H (similar to Greek catalexis, as discussed by Annis 2006). As a result, HLLH and LHLH shorten to HLH and LHH respectively. The fact that the meter changes midway in the poem (from variants of LHLH-LHLH to variants of LHLH-LHH) in cases like (10) is another deviation from Classical Persian metrics that is quite common in folk songs.

The high degree of metrical flexibility observed in these poems must not lead one to conclude that ‘anything goes’ in them. Note that the use of H in place of L is relatively rare and the use of L in place of H is non-existent. As for optional vowel lengthening (which allows us to treat many syllables as either L or H), it must be noted that it can only affect syllables with long vowels (roughly half the syllables) and even then the syllable cannot be read as L if it has a coda. The available readings for different syllable types in colloquial Persian poetry can be summarized as shown in Table 1.

**Table 1.** Available readings for syllable types in colloquial Persian poetry

Syllable structure	Metrical reading(s)
CV	L
CVC	H
CVCC	H (rarely as S, usually avoided altogether)
CVV	L, H
CVVC	H (rarely as S)
CVVCC	H (rarely as S, usually avoided altogether)

## 4. Alternative theories

In order to establish the claim that folk songs and pop lyrics follow quantitative meter with the properties described in the previous section, it seems necessary to compare the current proposal with the main competing theory. This theory involves stress, advocated to varying degrees by Khanlari (1948) and Adib Toosi (1953), and most recently by Tabibzadeh (2003, 2010), Tabibzadeh & Mirtalaei (2015), and Azarmakan & Nejati Jazeh (2014). Tabibzadeh (2003, 2010) states that stress (or pitch accent) in this sense does not necessarily match the predictable word-level stress patterns of Persian. Instead, he argues that stress can be assigned by the reader to any syllable she chooses to. According to him, for a line to follow a particular metrical pattern, the only necessary requirement is for the line to have exactly as many syllables as the desired metrical pattern; the stresses can be assigned by the reader in a way that matches the metrical pattern (Tabibzadeh 2010).

### 4.1 Falsifiability

The most important problem with Tabibzadeh's theory is that the stress-related component of his theory hardly seems to be falsifiable given that he assumes that the stresses can be assigned arbitrarily by the reader. In fact, even though my native judgments do not agree with Tabibzadeh's judgments on stress assignment during recital in folk songs, since the stress positions he advocates are not predictable by looking at the line, no poem can be pointed to as a counterexample to Tabibzadeh's theory. The only way the theory can be falsified, therefore, is acoustic analysis. He does not, however, mention the acoustic cues for the type of stress he is referring to.

The part of Tabibzadeh's theory that does indeed seem to be falsifiable is his claim about syllable count. In this case, however, there are many counterexamples to his theory. In fact, in all of the cases where H syllables are replaced with LL syllables (as happens in both (8) and (10)), syllable count between the lines does not match. Proponents of this theory do observe these discrepancies, but dismiss them by assuming such verses to underlyingly follow the desired syllable count (e.g., Ziamajidi & Tabibzadeh 2011). This approach, however, undermines the falsifiability of the theory again, since no systematic account that can predict the occurrence of unequal syllable counts is presented.

It is worth mentioning that as Tabibzadeh (2003) acknowledges, the metrical patterns that are argued for in the non-quantitative approach are mostly stress-based counterparts of the quantitative patterns of Classical Persian poetry, with stress replacing heavy positions and lack of stress replacing light positions. The difference between the two theories, therefore, is primarily in how the correspondences between the lines and their respective meters are established.

## 4.2 Poems claimed to lack quantitative meter

In § 3, various poems were shown to be analyzable through the presented quantitative approach. One could argue, however, that the particular poems that were examined were not necessarily of the type that proponents of the non-quantitative approach have in mind. To respond to this concern, a number of poems examined by Tabibzadeh & Mirtalae (2015) are analyzed here.

The poem in (11) is presented by Tabibzadeh & Mirtalae (2015) as a stress-based poem that lacks quantitative meter. Under the theory presented in this chapter, however, its analysis as a poem with quantitative meter is quite straightforward.

- (11) LLHL L H HL LH HL LHH  
*midunesti ke χak færf-e mæne: ræfti næmundi:*  
 you.knew that soil rug-of me-is you.left you.didn't.stay  
 "You knew that I have no rugs but the earth itself, and you left me."  
LL HL LHH  
*tjêra bæxt-e sefi:d-o:*  
 why luck-e white-OBJ  
 "My brilliant luck"  
 L LHL LHH  
*be sia:hi nefundi:*  
 to blackness you.made.sit  
 "Why did you make it black"  
LLHL LH L H LH HL LHH  
*midunesti fægæt to ro: daræm ræfti næmundi*  
 you.knew only you OBJ I.have you.left you.didn't.stay  
 "You knew that you are all I have, and you left me"  
LL HL LHH  
*tjêra morg-e omi:d-o:*  
 why bird-of hope-OBJ  
 "The bird of hope"  
 LH HL LHH  
*æz-in χu:ne pærunði*  
 from-this house you.made.fly  
 "Why did you make it fly away?"

In the word /æzin/ in the last line, a heavy syllable is parsed as light, similar to the case discussed in the previous section. The metrical patterns of the lines are fragmented with spaces and shown in (12), replacing the H of /æzin/ with an L.

- (12) LLHLL HHLL HHLL HH  
 LLHLL HH  
 LLHLL HH  
 LLHLL HHLL HHLL HH  
 LLHLL HH  
 LLHLL HH

The basic meter is the familiar meter HHLL-HHLL-HHLL-HH of Classical Persian poetry (belonging to the same meter family as the Ruba'ee meter), but the initial H is replaced with LL in all lines, similar to the cases discussed in the previous section. The fact that the verses have different numbers of HHLL but all end in HH is reminiscent of the modern style of Persian poetry known as *she'r e now*.

Another interesting example is a poem that is considered by Tabibzadeh & Mirtalae (2015) to have neither quantitative nor stress-based meter. The quantitative analysis of this poem under the current theory is shown in (13). Since the meter is a variant of LLHH-LLHH-LLH, the verse-initial syllable may appear as either L or H.

- (13) H LHH L HHL LHH LL HH  
*gæm miu:ne: dota tʃefmu:n-e gæfsæng-et lune kærde:*  
 sorrow between two eyes-e pretty-your nest has.made  
 “Sorrow has built a nest between your two beautiful eyes”  
 H L HHL LHH LL HH  
*fæb tu mu:ha:je sia:h-et χune kærde:*  
 night in hairs-e black-your house has.made  
 “The night has built a house inside your black hairs”  
 LL HHL LHH LL HHL LH  
*dota tʃefmu:n-e sia:h-et mese fæbha:-je mæn-e:*  
 two eyes-e black-your like nights-of me-is  
 “Your two dark eyes are like my nights”  
LLHHL L HH LL HHL LH  
*siahi:ha:-je do tʃefma-t mese gæmha:-je mæne:*  
 darknesses-of two eyes-your like sorrows-of me-is  
 “The darkness of your two eyes is like my sorrows”  
HL HH LLH LH LH LH LH  
*vægti bogz-æz mozeha-m pajin miad barun mijfe:*  
 when tears-from eyelashes-my down comes rain becomes  
 “It rains when tears begin to drop from my eyelashes”  
 HL H HLLH HLL HH  
*sejl-e gæm a:badi-m-o: viz:rune kærde:*  
 flood-of sorrow village-my-OBJ ruins has.made  
 “The flood of sorrow has brought my village to ruins”



HL H H LLH HLLH H LLH  
*vægti ba: mæn mimuni: tænhaji-m-o: bad mibære:*  
 when with me you.stay loneliness-my-OBJ wind carries  
 “When you stay with me, the wind carries away my loneliness”  
LL HH LLH LHL HH  
*dota t̂f̂f̂ma-m barun-e: sæbu:ne kærde:*  
 two eyes-my rain-e nightly has.done  
 “My two eyes are raining during the night”  
LLH HLL H H LL H  
*bæhar-æz dæstaje mæn pær zæd-o ræft*  
 spring-from handsof me feather hit-and went  
 “The spring flapped its wings and flew away from my hands”  
 LL H HL LH LHL HH  
*gol-e jæχ tu:je del-æm d̂zævu:ne kærde*  
 flower-of ice in heart-my sprout has.made  
 “Wintersweet flowers have blossomed in my heart”

The mappings for this poem are more straightforward, but the metrical structure needs more explanation. The metrical patterns of the lines are shown in (14). Verse-initial H's are replaced by L's.

- (14) a. LLHH LLHH LLHH LLHH  
 b. LLHH LLHH LLHH  
 c. LLHH LLHH LLHH LLH  
 d. LLHH LLHH LLHH LLH  
 e. LLHH LLHLHLHLHLH  
 f. LLHH LLHH LLHH  
 g. LLHH LLHH LLHH LLH  
 h. LLHH LLHLHLHH  
 i. LLHH LLHH LLH  
 j. LLHH LLHLHLHH

The underlying metrical structure seems to be repetitions of LLHH, followed in some of the lines by a catalectic version of it (LLH). In two of the lines, (14h) and (14j), however, we see the long sequence LLHLHLHH instead of LLHH-LLHH. In other words, as the underlined characters demonstrate, an HL sequence is replaced by an LH sequence in these lines, making them different from the other lines. This alternation is not attested in Classical Persian poetry (hence a difference, although small, between the two poetic traditions), but the pattern LLHLHLHH itself is a common pattern in Persian poetry (known as *Ramal Mashkool*), and an alternation between these meters as seen in (14) is in fact exactly the same alternation that is found in another quantitative system, i.e. Greek metrics, between Ionic dimeter

(LLHH LLHH) and anacreontic (LLHLHLHH), through a process of swapping (anacalasis) of the two syllable positions (Annis 2006). In line (14e), two swapping operations of this kind seem to be at play, resulting in LLHLHLHLHLH instead of LLHH-LLHH-LLH.

One may wonder why swapping is allowed only in certain positions, and what makes these positions special. This falls beyond the scope of this chapter, but it is worth noting that as I have shown in Mahdavi Mazdeh (2019: 165–169), the swapping observed here in LLHH-LLHH and the one observed in (10) in HLLH-HLLH are in fact instances of the same phenomenon. From a rhythmic perspective, the correct fragmentation for the former meter is LLH-HLLH-H (I refrain from representing the meters with those fragmentations here merely because most readers are familiar with the conventional fragmentations), which makes it clear that in both cases swapping targets HL sequences in the beginning of HLLH blocks.

Cases such as the one discussed above show that a quantitative approach to Persian folk songs and pop lyrics not only provides a better account of the metrical system of these poems, but by introducing a new class of metrical poems that are closely related to Classical Persian poetry opens the way for more in-depth research on Persian metrics and quantitative meters in general.

## 5. Corpus data

I now present results of a corpus analysis examining how successful this chapter's proposal is in accounting for available data. I start with pop lyrics, which generally tend to follow the metrical rules more steadfastly. Table 2 shows the meters used in sample poems from five well-known contemporary lyricists (Ardalan Sarfaraz, Iraj Jannati Atayi, Yaghma Golrooyi, Maryam Heydarzadeh, and Shahyar Ghanbari). As the numbers in the table show, there is usually less than one case of optional vowel shortening occurring per verse (*mesra*). There are many cases of swapping adjacent L and H syllables in certain positions of certain meters, and merely 5 cases of H being used in an L position. The other type of poetic license mentioned earlier, i.e. using LL in place of H, is not used in any of the poems examined. Moreover, the lengthening of a short vowel, as expected, is non-existent. There is also one case of using a superheavy syllable in place of an H in the poem *madar-bozog kojayi* by Golrooyi, which is not reflected in the table. Note that in some cases only the first few verses of the poem have been examined. In the table, OVS denotes 'Optional Vowel Shortening'. The cases where a verse-initial L appears as H in meters that allow this in classical Persian poetry (as discussed earlier) are not included in this table.

In poems where swapping occurs, the consecutive syllables for which swapping may occur are underlined in the “Meter” column.

**Table 2.** Poetic license in pop lyrics

Poem	Meter	Verses	OVS	Swap	H for L
Gharib-e ashena (Sarfaraz)	LHHH LHHH LHHH LH (after 2: LHH HLH)	14	8		
Man o to (Sarfaraz)	LHHH LHHH LHHH LHH	16	16		
Do panjere (Sarfaraz)	LL <u>HH</u> <u>LL</u> HH	32	25	15	
Ghoruba ghashangan (Sarfaraz)	LL <u>HH</u> <u>LL</u> HH <u>LL</u> HH LLH (verses 11,12: LHLH LHLH)	14	15	4	
Razeghi (Jannati)	LL <u>HH</u> <u>LL</u> HH (verses 7 to 10: HLHH HLHH H)	22	10	5	
Ye nafar ye ruz miad (Jannati)	LL <u>HH</u> <u>LL</u> HH <u>LL</u> H	24	23	15	
Nemikham mesl-e hame gerye konam (Jannati)	LL <u>HH</u> <u>LL</u> HH <u>LL</u> H	18	13	15	
Shab-e shishe'i (Jannati)	LL <u>HH</u> <u>LL</u> HH <u>LL</u> H	18	20	9	
Eshgh-e man ashegham bash (Jannati)	<u>HLLH</u> <u>LHH</u>	36	9	21	1 (v. 17)
Baraye ebrat-e in e (Golrooyi)	LL <u>HH</u> <u>LL</u> HH (verses 17 to 20: <u>HLLH</u> <u>LHLH</u> )	52	47	18	1 (v. 17)
Madarbozorg kojayi (Golrooyi)	<u>HLLH</u> LHH	16	11	6	
Barpa (Golrooyi)	<u>HLLH</u> <u>LHLH</u>	16	20	9	
Espania espania (Golrooyi)	LL <u>HH</u> <u>LL</u> HH	8	1	3	
Chehar rah (Golrooyi)	LL <u>HH</u> <u>LL</u> HH <u>LL</u> HH <u>LL</u> H	8	11	5	
Mesl-e hichkas (Heydarzadeh)	LLHH LLHH	20	12		
In ruza (Heydarzadeh)	<u>HLLH</u> <u>HLLH</u>	16	11	8	1 (v. 14)
Name-ye bijavab (Heydarzadeh)	<u>HLLH</u> <u>HLLH</u>	12	9	12	1 (v. 11)
Boo-ye gandom mal-e man (Ghanbari)	LLHH LLHH LLHH LLH	6	11	2	
Hamishe ghayeb (Ghanbari)	LL <u>HH</u> <u>LL</u> HH <u>LL</u> HH <u>LL</u> H	6	5	4	
Ghesse-ye barre o gorg (Ghanbari)	LL <u>HH</u> <u>LL</u> H	16	13	6	1 (v. 6)

As the data in Table 2 show, lyricists are generally observant of metrical requirements. In contrast, violating metrical requirements is relatively common in folk songs. Table 3 shows data from folk songs. The songs are the first 10 songs in the list provided by Tabibzadeh (2003), skipping those that are composed by well-known poets. As before, for some of the poems only the first few verses are examined. In each of the first and the second to last poems (marked with asterisks), there is one case of swapping between the two final syllables of HLLH, which is not reflected in the table.

Table 3. Poetic license in folk songs

Poem	Meter	Verses	OVS	Swap	H for L	LL for H
Usta baba bozi dasht	<u>HLLH</u> LHH*	10	14	4	2	3
Kalaghe mige man gharoghar mikonam vasat	HLLH HLLH HLLH HLLH HLLH	2	21		6	7
Gonjishkak-e alili	HLLH LHH	5	8	3	1	1
Ozra kxanom salam-o alek ya allah	<u>HLLH</u> HLLH LHH	8		2		3
In kie taptap mikone	HLLH HLLH	2	6			1
Atish darin balatarak	<u>HLLH</u> <u>HLLH</u>	5	13	3	1	5
Yal-e man yaragh mikhad	(LLH) <u>LHLH</u> <u>LHLH</u>	6	18	1	2	2
Mardi ke nun nadare	HLLH LHH	4	4		1	2
Ey khoda sukhte junam	LH <u>HLLH</u> H*	7	3	1		
Harki be fekr-e khish e	<u>HLLH</u> LHH	4	2	1		

The main interesting issue in these songs is the use of an H in place of an L (second to last column). Such cases are rare in Table 2 but relatively common in some of the rows of Table 3. It is also worth noting that optional vowel shortening occurs much more frequently in these poems. In fact, it seems that the default reading of a traditionally long vowel is short, suggesting that, even within colloquial Persian, vowel length distinctions are less prominent in less formal contexts.

## 6. Conclusion

In this chapter, I have argued that Persian folk songs and pop lyrics have quantitative meter, based on the same metrical assessment process as Classical Persian poetry and differing from it only in the way lines are mapped to metrical patterns (the correspondence criteria). Moreover, I introduced the phenomenon of ‘Optional Vowel Shortening’ that is specific to the correspondence criteria of folk

songs and pop lyrics, and accounts for the vast majority of its differences from Classical Persian poetry.

In addition to optional vowel shortening, I pointed to a number of less important innovations in the metrical system of these songs including more extensive use of LL in place of an H, more extensive use of LHLH in place of HLLH, and a process of swapping that is reminiscent of anacalasis in Greek metrics.

I also discussed briefly the possibility that optional vowel shortening is a reflection of the phonological properties of colloquial Persian as opposed to formal Persian, the language of Classical Persian poetry. This proposal may shed light on the phonology of other languages (such as Gilaki) that have been mentioned in the literature to follow a metrical system similar to that of Persian folk songs. If the current research is on the right track, it can open new doors to the study of both the phonology of these languages and the study of quantitative metrics in general.

## References

- Adib Toosi, Mohammad Amin. 1953. Tarānehā ye mahalli [Folk songs]. *Journal of the Department of Literature and Humanities of Tabriz University* 5(1). 49–101. <http://www.ensani.ir/fa/content/160527/default.aspx>
- Annis, William S. 2006. Introduction to Greek meter. Aoidoi.org. <http://www.aoidoi.org/articles/meter/intro.pdf> (29 November, 2017).
- Azarmakan, Heshmatollah & Morteza Nejati Jazeh. 2014. Vazn e she'r e kudak: aruzi? hejayi? ya tekye'i-hejāyi? [Meters in children's songs: Quantitative, syllabic, or stress-syllabic?]. *Adab Pajouhi* 8 (4). 101–116. [http://adab.guilan.ac.ir/article\\_38.html](http://adab.guilan.ac.ir/article_38.html)
- Deo, Ashwini & Paul Kiparsky. 2011. Poetries in contact: Arabic, Persian, and Urdu. *Linguistic Insights – Studies in Language and Communication* 113. 145–172.
- Ehterami, Manoochehr. 1987. *Hasani e mā ye barre dāsht*. Tehran: Lakposht.
- Etesami, Parvin. In *ghat'e rā barāye sang e mazār e khodam sorude'am*. Ganjoor online poetry collection. <https://ganjoor.net/parvin/divanp/mtm/sh166/> (29 November, 2017).
- Ghanbari, Shahyar. 1998. *Daryā dar man* [The sea in me]. Tehran: Javidan Publications.
- Hayes, Bruce. 1979. The rhythmic structure of Persian verse. *Edebiyat: Journal of ME Literatures* 11. 193–242.
- Khanlari, Parviz Natel. 1948. *Tāghigh e enteghādi dar aruz e fārsi va chegunegi e tahavvol e owzān e ghazal* [Critical analysis in Persian metrics and the developments of meters in ghazal]. Tehran: University of Tehran Press.
- Mahdavi Mazdeh, Mohsen. 2019. The rhythmic structure of Persian poetic meters. Tucson, AZ: University of Arizona dissertation.
- Najafi, Abolhassan. 2015. *Darbāreye Tabaghebandi e vaznhā ye she'r e fārsi* [On classifying Persian poetic meters]. Tehran: Niloofar.
- Tabibzadeh, Omid. 2003. *Tahlil e vazn e she'r e āmiāne ye fārsi* [Analysis of meter in Persian folk songs]. Tehran: Niloofar.

- Tabibzadeh, Omid. 2010. Barrasi e tatbighi e vaznhā ye kammi va tekye' i hejāyi dar fārsi va gilaki [A comparative study of quantitative and accentual-syllabic meters in Fārsi and Gilaki]. *Adab Pajouhi* 4(1). 7–30. [http://adab.guilan.ac.ir/article\\_362.html](http://adab.guilan.ac.ir/article_362.html)
- Tabibzadeh, Omid & Maedeh Mirtalae. 2015. Vijhegihā ye vazni dar tarānehā va tasnifhā ye fārsi [Metrical properties of Persian song lyrics]. *Nameh Ye Farhangestan* 14(3). <http://www.ensani.ir/fa/content/345768/default.aspx>
- Thiesen, Finn. 1982. *A manual of classical Persian prosody: With chapters on Urdu, Karakhanidic and Ottoman prosody*. Wiesbaden: Harrassowitz.
- Toosarvandani, Maziar Doustdar. 2004. Vowel length in modern Farsi. *Journal of the Royal Asiatic Society* 14(3). 241–251. <https://doi.org/10.1017/S1356186304004079>
- Shamisa, Siroos. 2004. *Āshnāyi bā aruz va ghāfieh* [An introduction to meter and rhyme]. Tehran: Mitra.
- Vahidian Kamyar, Taghi. 1979. *Barrasi e vazn e she'r e āmiāne ye fārsi* [A review of meters in Persian folk songs]. Tehran: Agah.
- Ziamajidi, Leyla & Omid Tabibzadeh. 2011. Barrasi e vazn e she'r e āmiāne ye fārsi bar asās e nazarie ye vazni [A review of meters in Persian folk songs based on metrical theory]. *Adab Pajouhi* 5(2). 59–79. <http://www.sid.ir/Fa/Journal/ViewPaper.aspx?id=153416>



# Stripping structures with negation in Persian

Vahideh Rasekhi

University of California, Los Angeles

In this chapter, I introduce two novel Stripping constructions from Persian that occur with negation. I refer to these constructions as Polarity Stripping and Negative Stripping. I argue that they involve clausal coordination, and that in the second coordinate, the entire clause except for a constituent is elided under identity with corresponding parts of the first coordinate. I propose that these constructions involve TP ellipsis, which is licensed by the Pol head that carries an [E] feature (Merchant 2001). I also study another structure, which I refer to as Pseudo-stripping. I argue that even though Pseudo-stripping looks like Polarity Stripping and Negative Stripping, it does not involve ellipsis despite what has been claimed in the literature for English (Kolokonte 2008). I propose that Pseudo-stripping is mono-clausal and is derived via movement.

**Keywords:** Ellipsis, Polarity Stripping, Negative Stripping, Pseudo-stripping, Persian

## 1. Introduction

The goal of this chapter is to provide syntactic analyses for Stripping structures with negation in Persian, which have not been previously studied. Stripping refers to an operation in which the entire clause except for one constituent is elided under identity with corresponding parts of the preceding clause (1).

(1) *Alan likes to play volleyball, but not Sandy.*

(Hankamer & Sag 1976: Example 44)

In the second conjunct in (1), the entire clause is elided except for one element *Sandy*, which is referred to as *remnant*, and the negative element *not*.<sup>1</sup>

---

1. This element is not limited to negation but rather it can also be an affirmative element or an adverb. Here we focus only on structures that include a negative marker. For discussion on different types of stripping in Persian refer to Rasekhi (2018, 2019a, 2019b 2019c).



Consider the Persian Stripping constructions in (2). In both (2a) and (2b), the remnant *Ayda*, which precedes the negative marker *na*, carries a high pitch accent.<sup>2</sup> In addition, it contrasts with its corresponding element *Araz* in the preceding clause. The only difference between these structures, on the surface, is the presence of the coordinator *vali* “but” in (2a). I refer to the structure in (2a) as Polarity Stripping (PolS) and to the structure in (2b) as Negative Stripping (NegS).

- (2) a. ARAZ ketāb kharid,      **vali** **AYDA** **na**                      (PolS)  
           Araz book bought.3SG but Ayda NEG  
           “Araz bought books, but Ayda did not (buy books).”  
       b. ARAZ ketāb kharid,      **AYDA** **na**                      (NegS)  
           Araz book bought.3SG Ayda NEG  
           “Araz bought books, Ayda did not (buy books).”

Persian also allows structures such as (3), which I refer to as Pseudo-Stripping (PseS). In this construction, unlike PolS and NegS, the negative marker carries a high pitch accent.

- (3) ARAZ ketāb kharid,      **NA** **Ayda**                      (PseS)  
           Araz book bought.3SG NEG Ayda  
           “Araz bought books, not Ayda.”

On the surface, the difference between the structures in (2) and (3) is in the order of *Ayda* and the negative marker *na*. However, I show that they are different structures and cannot be accounted for in the same way. I argue that PolS and NegS are true cases of Stripping, which involve ellipsis. On the other hand, based on the evidence from Persian, I argue that Pseudo-stripping does not involve ellipsis despite what has been claimed for English (Kolokonte 2008).

Adopting Rizzi’s (1997) cartographic approach, I propose that the remnant in PolS and NegS moves to the Spec of TopP and FocP, respectively. In addition, I propose that the negative marker *na* originates in the Spec of PolP, and the Pol head, which carries an [E] feature (Merchant 2001), licenses the deletion of its complement, TP, at the PF level.

On the other hand, I propose that PseS is derived via movement rather than ellipsis. I propose that the negative marker in this structure is a constituent negation, and that the NEG XP<sup>3</sup> constituent, in the underlying structure, adjoins to XP’s corresponding element. However, since it carries a contrastive focus feature, it has

---

2. The remnant in these constructions can be a subject, an object, an adverb, or an adjective. However, in this chapter, for reasons of space, the presented data only includes subject remnant.

3. XP refers to any element (e.g. DP, PP, AdvP, AdjP) that can follow the negative marker.

to move to a focus position. I propose that it undergoes rightward movement and adjoins to FocP in the TP level.

Structure of this chapter is as follows: In § 2, I discuss the differences among PolS, NegS, and PseS in terms of their context of occurrence and interpretation. In § 3, I provide evidence that the XP in PolS is a topicalized element while the XP in NegS and PseS is a focalized element. In § 4, I discuss the nature of the negative marker in these constructions. In § 5, I provide syntactic analyses for the constructions under discussion and § 6 provides a conclusion.

## 2. Context of occurrence and interpretation

In this section, I show that PolS, NegS, and PseS constructions cannot occur in the same contexts since they have different interpretations. To illustrate their differences, suppose that two friends A and B are talking to each other about the shopping they did with their friends. Without any further discourse, speaker A can utter the sentence in (4a), which is an instance of PolS. However, as the ungrammaticality of the sentences in (4b) and (4c) shows, NegS and PseS are not possible without a linguistic antecedent.

- (4) Context: [A and B are talking about the shopping they did with their friends]
- a. *ARAZ dirooz ketāb kharid, vali AYDA na* (PolS)  
 Araz yesterday book bought.3SG but Ayda NEG  
 ‘‘Araz bought books yesterday, but Ayda did not (buy books).’’
- b. \**ARAZ dirooz ketāb kharid, AYDA na* (NegS)  
 Araz yesterday book bought.3SG Ayda NEG
- c. \**ARAZ dirooz ketāb kharid, NA Ayda*  
 Araz yesterday book bought.3SG NEG Ayda

The grammaticality of (4a) shows that PolS can occur in an out of the blue context, and that it does not require a linguistic antecedent. Therefore, we can say that in this structure, the speaker is introducing new information to the discourse by saying that *Araz bought books but Ayda did not buy books*. On the other hand, the ungrammaticality of (4b) and (4c) suggests that NegS and PseS cannot occur in an out of the blue context, which is due to the fact that these constructions have an obligatory corrective interpretation. We expect these constructions to be acceptable with a linguistic antecedent. This is borne out as shown in (5) and (6).

In (5a), speaker A makes an assertion that *Araz and Ayda bought books*. However, speaker B corrects speaker A by uttering the sentence in (5b), which is an instance of NegS. The fact that this structure is acceptable with a linguistic antecedent, as shown in (5b), but not with a contextual antecedent, as in (4b), shows that it can only occur in contexts in which a proposition has been made.

- (5) a. *dirooz Araz va Ayda ketāb kharid-an*  
 yesterday Araz and Ayda book bought-3PL  
 “Yesterday, Araz and Ayda bought books.”
- b. *ARAZ dirooz ketāb kharid, AYDA na* (NegS)  
 Araz yesterday book bought.3SG Ayda NEG  
 “(No, you are wrong), ARAZ bought books yesterday, AYDA did not (buy books).”

In (5b), the emphasis is on the predicate and whether *Ayda* and *Araz* bought books. We know that we have two alternatives in this regard; *Ayda* and *Araz* have either bought books or they have not bought books. Therefore, we can say that in NegS, as in (5b), speaker B corrects speaker A’s statement by showing which alternative is true and which one is false. In this context, *Araz bought books* is true while *Ayda bought books* is false.

PseS, similar to NegS, is acceptable when it is provided with a linguistic antecedent (6b).

- (6) a. *dirooz Ayda ketāb kharid*  
 yesterday Ayda book bought.3SG  
 “Yesterday, Ayda bought books.”
- b. *ARAZ ketāb kharid, NA Ayda* (PseS)  
 Araz book bought.3SG NEG Ayda  
 “(No, you are wrong), ARAZ bought books, NOT Ayda.”

In (6a), speaker A makes an assertion that *Ayda bought books*. However, speaker B corrects speaker A’s assertion by saying (*no, you are wrong*) *it was Araz who bought books, not Ayda*. In (6b), the emphasis is on the person who bought books; whether it was *Araz* or *Ayda* who bought books. We can say that in this context, we have two alternatives with regard to who bought books; either *Araz bought books* or *Ayda bought books*. We see that speaker B in (6b) corrects speaker A by replacing *Ayda* in (6a) with *Araz*. Thus, we can say that in Pseudo-stripping, we have correction by substitution.

### Summary

A summary of the characteristics of PolS, NegS, and PseS is presented in (7).

(7) Context of occurrence and interpretation	PolS	NegS	PseS
a. Acceptable in an out of the blue context	Yes	No	No
b. Has a corrective interpretation	No	Yes	Yes
c. Has correction by showing which alternative is true and which one is false	NA	Yes	No
d. Has correction by substituting one alternative by another	NA	No	Yes

### 3. Information structure

In this section, I discuss how PolS, NegS, and PseS structures are constrained by information structure. I argue that the XP in PolS has the characteristics of a topic while the XP in NegS and PseS has the characteristics of focus. I provide two pieces of evidence for this claim.

One difference between topic and focus is that we can have more than one topic in a sentence while we can only have one focus per sentence.<sup>4</sup> With this in mind, let us have a look at the structures in (8)–(10).

- (8) *AYDA ketāb-a-ro kharid, vali ARAZ majalle-a-ro na*  
 Ayda book-DEF-ACC bought.3SG but Araz magazine-DEF-ACC NEG  
 “Ayda bought the book but Araz didn’t (buy) the magazine.” (PolS)
- (9) \**AYDA ketāb-a-ro kharid, ARAZ majalle-a-ro na* (NegS)  
 Ayda book-DEF-ACC bought.3SG Araz magazine- DEF-ACC NEG
- (10) \**AYDA ketāb-a-ro kharid, NA Araz majalle-a-ro* (PseS)  
 Ayda book-DEF-ACC bought.3SG NEG Araz magazine- DEF-ACC

In PolS (8), we have two elements before the negative marker; the subject *Araz* and the direct object ‘magazine’, and the sentence is grammatical. However, as the ungrammaticality of the sentences in (9) and (10) shows, it is not possible to have two elements in NegS and PseS. These examples suggest that PolS is compatible with topicalized elements while NegS and PseS are compatible with focalized elements.

The second piece of evidence that indicates NegS and PseS are compatible with focalized elements while PolS is not comes from their compatibility with focus adverbs such as *only*. Let us consider the examples in (11).

- (11) [Context: speaker A and speaker B are talking about their friends who bought books]  
 A. *ki-yā ketāb kharid-an?*  
 who-PL book bought-3PL  
 “Who bought books?”  
 B. \**hame ketāb kharid-an, vali FAQAT Ayda na* (PolS)  
 everyone book bought-3PL but only Ayda NEG  
 Intended: “Everyone bought books, but only Ayda did not.”  
 (Adapted from López & Winkler 2000)

4. Persian allows two elements bearing a contrastive focus feature in the same sentence only if one of them has an inherent focus feature (Karimi 2005: 133).

- (i) *KIMEA māh-e gozashte faqat se-tā film did-e*  
 Kimea month-EZ previous only three-part film saw-3SG  
 “It was Kimea who saw only three movies last month.” (Everyone else has seen more movies)

The ungrammaticality of the sentence in (11b) shows that the DP *Ayda* in PolS is not compatible with the focus adverb *faqat* ‘only’. In the same context, the conversation can continue as in (12). In (12a), the speaker makes an assertion that *only Araz bought books*. We see that NegS (12b) and PseS (12b’) are acceptable in this context, which shows that the DP *Araz* in these constructions is compatible with the focus adverb ‘only’.

- (12) a. *pas faqat Araz ketāb kharid*  
 so only Araz book bought.3SG  
 “So, only Araz bought books.”
- b. *HAME ketāb kharid-an, FAQAT Araz na* (NegS)  
 everyone book bought.3PL only Araz NEG  
 “(No, you are wrong) everyone bought books and not just Araz.”
- b'. *HAME ketāb kharid-an, NA FAQAT Araz* (PseS)  
 everyone book bought.3PL NEG only Araz  
 “(No, you are wrong), everyone bought books, not just Araz.”
- (Adapted from López & Winkler 2000)

Now that we have established we have a topicalized element in PolS but a focalized element in NegS and PseS structures, we need to discuss the nature of the negative marker in these structures.

#### 4. The nature of negative marker

In PolS (13a) and NegS (13b), when the sentence is continued, we have two negative markers in the second clause: one is after the remnant *Araz* and the other is affixed to the verb. However, in PseS (13c), we have only one negative marker that precedes *Araz* and it is not possible to continue the sentence after *Araz*.

- (13) a. *AYDA ketāb kharid, vali ARAZ na, ketāb na-kharid*  
 Ayda book bought.3SG but Araz NEG book NEG-bought.3SG  
 “Ayda bought a book, but Araz did not buy a book.”
- b. *AYDA ketāb kharid, ARAZ na, ketāb na-kharid*  
 Ayda book bought.3SG Araz NEG book NEG-bought.3SG  
 “Ayda bought a book, Araz did not buy a book.”
- c. *AYDA ketāb kharid, NA Araz (\*ketāb na-kharid)*  
 Ayda book bought.3SG NEG Araz book NEG-bought.3SG  
 “Ayda bought a book, not Araz.”

The data in (13) indicates that: i) PolS and NegS are bi-clausal while PseS is mon-clausal, ii) in PolS (13a) and NegS (13b), we have two negative markers in the underlying structure while in PseS (13c), we have only one negative marker. The question that now needs to be addressed is: What is the nature of the negation in these constructions?

In languages like English, in which the sentential negation ‘not’ and constituent negation ‘no’ are different, only the sentential negation is possible in Stripping, as shown in (14a), and phrasal negation is not possible (14b).

- (14) a. *John bought a book, not Mary.*  
 b. \**John bought a book, no Mary.*

However, in Persian, the sentential negation (15) and phrasal negation (16) are homophonous. Thus, it is not clear whether the negative marker in the constructions in (13) is sentential or constituent negation.

- (15) Q. *Maryam-ro did-i?*  
*Maryam-ACC saw-2SG*  
 “Did you see Maryam?”

- A. *na, na-did-am*  
 NEG NEG-saw-1SG  
 “No, I did not see her.”

- (16) a. *na man chini sohbat mikon-am, na unā*  
 NEG I Chinese harf do-1SG NEG they  
 “Neither I nor they speak Chinese.”

- b. *man ketāb mikhoon-am, na majale*  
 I book read-1SG NEG magazine  
 “I am reading a book, not a magazine.”

(Kwak 2010: 624)

I propose that the negative marker in PseS is constituent negation. This claim is based on Klima’s (1964) either-and neither conjoining test and adverbs, which I discuss in the next two subsections. Then, I propose that the negative marker in PolS and NegS functions as a focusing adverb.

#### 4.1 Either-and neither conjoining test

In this section, I use Klima's (1964) either-and neither-conjoining test to determine the nature of negative marker in PolS, NegS, and PseS. As shown in (17a) and (18a), the structures with sentential negation are grammatical with either-and neither-conjunction. However, the structures with constituent negation are not compatible with this type of conjunction, as shown in (17b) and (18b).

- (17) a. Sentential negation: *Mary isn't a happy person and John isn't either.*  
 b. Constituent negation: \**Mary is a not happy person and John isn't either.*
- (18) a. Sentential negation: *Mary isn't a happy person and neither is John.*  
 b. Constituent negation: \**Mary is a not happy person and neither is John.*

Let us apply this test to our constructions under discussion to determine whether the negative marker *na* in these structures is sentential negation or constituent negation. As the examples below show, we see that PolS (19) and NegS (20) constructions are acceptable with either-conjoining while PseS (21) is not. This shows that the negative marker in PseS is constituent negation.

- (19) *AYDA khoshhāl-e, vali ARAZ na, MARYAM ham na* (PolS)  
 Ayda happy-3SG but Araz NEG Maryam also NEG  
 "Ayda is happy but Araz is not (happy), Maryam is neither."
- (20) *AYDA khoshhāl-e, ARAZ na, MARYAM ham na* (NegS)  
 Ayda happy-3SG Araz NEG Maryam also NEG  
 "Ayda is happy, Araz is not (happy), Maryam is neither."
- (21) \**AYDA khoshhāl-e, NA Araz, NA ham Maryam* (PseS)  
 Ayda happy-3SG NEG Araz NEG also Maryam  
 [Intended meaning] "Ayda is happy, not Araz, neither is Maryam."

#### 4.2 Adverbs

The second piece of evidence that shows the negative marker in PseS, but not in PolS and NegS, is constituent negation comes from examples as in (22)–(24).

- (22) *AYDA hamishe ketāb mikhar-e, vali ARAZ hamishe na* (PolS)  
 Ayda always book buy-3SG but Araz always NEG  
 "Ayda always buys books, but Araz does not always (buy books)."
- (23) *AYDA hamishe ketāb mikhar-e, ARAZ hamishe na* (NegS)  
 Ayda always book buy-3SG Araz always NEG  
 "Ayda always buys books, Araz does not always (buy books)."
- (24) \**AYDA hamishe ketāb mikhar-e, NA hamishe Araz* (PseS)  
 Ayda always book buy-3SG NEG always Araz

As illustrated in these examples, in PolS (22) and NegS (23), an adverb can occur between *Araz* and the negative marker. However, this is not possible in PseS (24), which shows that the negative marker in this structure is constituent negation. Now that we know the negative marker in PseS involves constituent negation, in the next section, I show that the negative marker in PolS and NegS functions as a focusing adverb.

### 4.3 Focusing adverb

In this section, I propose that the negative marker in PolS and NegS, but not in PseS, belongs to a class of adverbs such as *never* called focusing adverb (Rooth 1985, 1996). This proposal is supported by examples such as (25)–(27).

- (25) AYDA *hamishe ketāb mi-khar-e*,    *vali* ARAZ *hichvaqt (ketāb*  
 Ayda always book DUR-buy-3SG but Araz never book  
*ne-mi-khar-e)*

NEG-DUR-buy-3SG

“Ayda always buys books, but Araz never (buys books).”

- (26) AYDA *hamishe ketāb mi-khar-e*,    ARAZ *hichvaqt (ketāb*  
 Ayda always book DUR-buy-3SG Araz never book  
*ne-mi-khar-e)*

NEG-DUR-buy-3SG

“AYDA always buys books, ARAZ never (buys books).”

- (27) \*AYDA *hamishe ketāb mi-khar-e*,    HICHVAQT Araz  
 Ayda always book DUR-buy-3SG never        Araz

In PolS (25) and NegS (26), it is possible to replace the negative marker with the adverb *never*, while in PseS (27), this is not possible. These examples show that the negative marker in PolS and NegS functions as a focusing adverb.

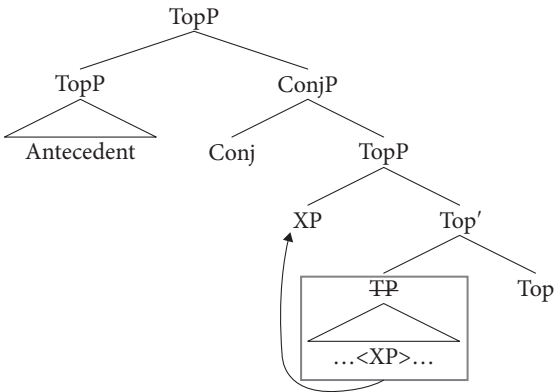
## 5. Analysis

### 5.1 Analysis of Polarity Stripping and Negative Stripping

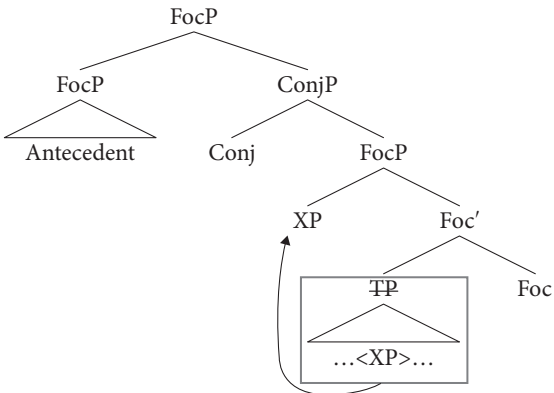
As discussed in § 3, we know that the remnants in PolS and NegS function as topicalized and focalized elements, respectively. I propose that the remnant in these structures carries contrastive topic and contrastive focus features; therefore, it moves to the Spec of TopP and to the Spec of FocP, respectively, as schematically illustrated in (28).



## (28) a. Structure of Polarity Stripping



## b. Structure of Negative Stripping



Regarding the negative marker, as discussed in § 3.5. I propose that the negative marker in these structures is a focusing adverb. In Persian, the position of sentential negation is assumed to be in the CP level, as the phrase structure in (29) illustrates.

(29) [CP [TopP [FocP [NegP [TP [ $\nu$ P PredP]]]]]] (Karimi 2005: 147)

Sentential negation, NegP, selects TP as its complement. However, negation is morphologically realized on the verb through Agree relation between Neg, which bears an interpretable negation feature, and  $\nu$ , which bears an uninterpretable negation feature (Taleghani 2008).

Since the negative marker in PolS and NegS precedes the sentential negation, which is suffixed on the verb, it has to be in a position higher than NegP. In addition, since the remnant in these constructions moves to the Spec of TopP and FocP,

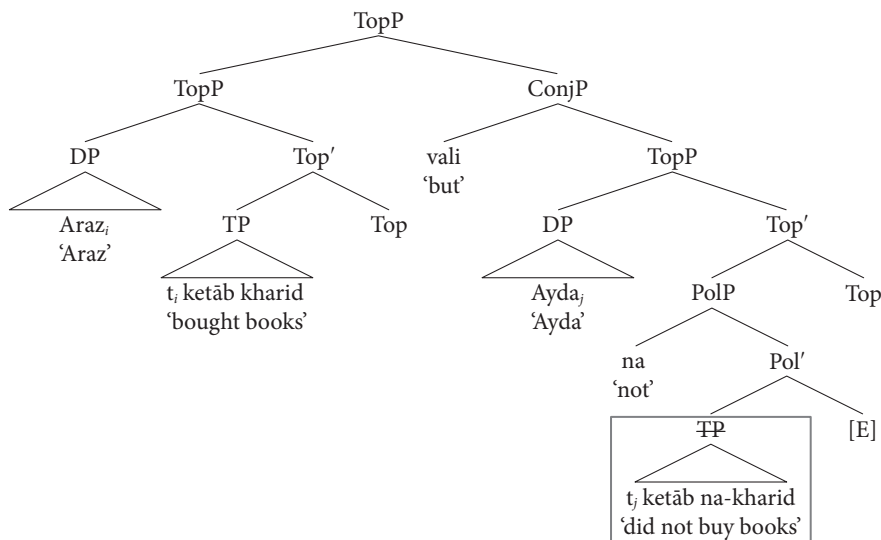
respectively, we know that the negative marker is higher than NegP but lower than FocP. Let us call this position PolP. If our analysis is on the right track, the phrase structure in (29) should be revised to include PolP.

(30) [CP [TopP [FocP [PolP [NegP [TP [vP PredP]]]]]]]

I propose that the negative marker *na* in PolS and NegS, which acts as a focusing adverb, originates in the Spec of PolP. I also propose that the Pol head carries an [E] feature (Merchant 2001) that licenses the deletion of its complement, TP, which is identical to the TP of the antecedent clause and therefore it becomes redundant, at the PF level.

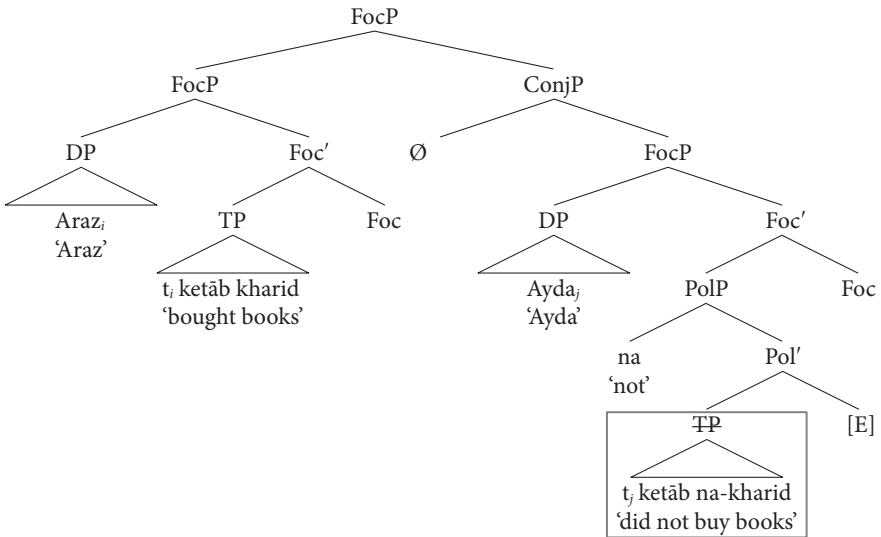
The structure I propose for PolS is given in (31). In this structure, the remnant *Ayda* in the second clause moves to the Spec of TopP and the [E] feature on the Pol head licenses the deletion of TP at the PF level.

(31) Structure of Polarity Stripping



Similarly, the structure of NegS is schematically represented in (32). Even though there is not an overt coordinator in this structure, I assume that we have a null conjunction in this structure. As shown in (32), the remnant *Ayda* in the second clause moves to the Spec of FocP, before TP deletion takes place.

## (32) Structure of Negative Stripping



## 5.2 Analysis of Pseudo-stripping

We already know that PseS is mono-clausal. In addition, as discussed in §§ 3–4, we know that the XP in this structure functions as a focalized element, and the negative marker is constituent negation.

In this section, I propose that PseS in Persian is derived via movement. This is different from what has been proposed for English, where Kolokonte (2008) has proposed that PseS involves TP ellipsis.

I propose that the underlying structure of PseS is as shown in (33). The *na Araz* constituent originates adjacent to its corresponding element *Ayda*. However, since this word order is not possible in Persian, the constituent *na Araz* undergoes obligatory rightward movement, as shown in (34).

(33) \*AYDA, NA Araz, ketāb kharid (PseS)

Ayda NEG Araz book bought.3SG  
[lit. "Ayda, not Araz, bought books."]

(34) AYDA t<sub>i</sub> ketāb kharid, [NA Araz]<sub>i</sub>  
Ayda book bought.3SG NEG Araz

A piece of evidence for the movement comes from case marking, as shown in (35). The DP *majale* "magazine" and its corresponding element *ketāb* "book" have *-ro* marking. If *na majale* were base-generated, we would not expect *majale* to have *-ro* marking.

- (35) *KETAB-RO khoond-am, NA majala-ro*  
 book-ACC read-1SG NEG magazine-ACC  
 [*lit.* “The book I read, not the magazine.”]

Even though the structure in (33) does not sound natural in Persian, the proposal that the NEG XP originates adjacent to the XP’s corresponding element and arrives at its surface position via movement is supported by the data from German (36) and Spanish (37). As illustrated in the following examples, the NEG XP can appear at the sentence-final position or in the middle of the sentence.

- (36) German  
 a. *Anna spielt Klavier, nicht Maria*  
 Anna plays piano not Maria  
 “Anna plays piano but not Maria.”  
 b. *Anna, nicht Maria, spielt Klavier*  
 Anna not Maria plays piano  
 “Anna, not Maria, plays piano.” (Thomas Graf p.c.)
- (37) Spanish  
 a. *Anna toca el piano, no María*  
 Anna plays the piano not Maria  
 “Anna plays piano, not Maria.”  
 b. *Anna, no María, toca el piano*  
 Anna not Maria plays the piano  
 “Anna plays piano, not Maria.” (José Elías-Ulloa p.c.)

The questions that need to be addressed are: How does the movement in (34) work? Where does the NEG XP move to? I propose that the *na Ayda* constituent undergoes right movement<sup>5</sup> and adjoins to the FocP in the TP level,<sup>6</sup> as schematically illustrated in (38). The DP moves to FocP to satisfy the uninterruptable strong focus feature [*uConf\**] that is on the Focus head.

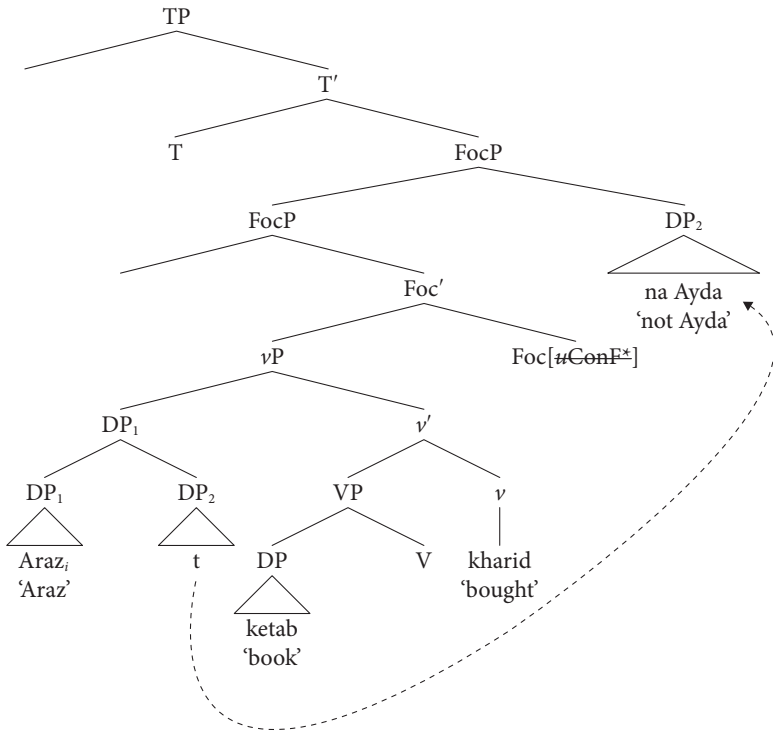
---

5. To derive PseS via leftward movement, one has to propose that the direct object ‘book’ and the verb ‘bought’ in (i) move out of the *vP*. This movement out of *vP* is not well-motivated.

(i) [<sub>FocP</sub> ARAZ<sub>i</sub> [<sub>TP</sub> ketāb<sub>j</sub> kharid<sub>k</sub> NA Ayda [<sub>vP</sub> t<sub>i</sub> t<sub>j</sub> t<sub>k</sub>]]] t<sub>i</sub> t<sub>j</sub> t<sub>k</sub>]]  
 Araz book bought.3SG NEG Ayda

6. See Kahnemuyipour (2001) and Rasekhi (2018) for the evidence on the existence of FocP in the TP level.

## (38) Structure of Pseudo-stripping



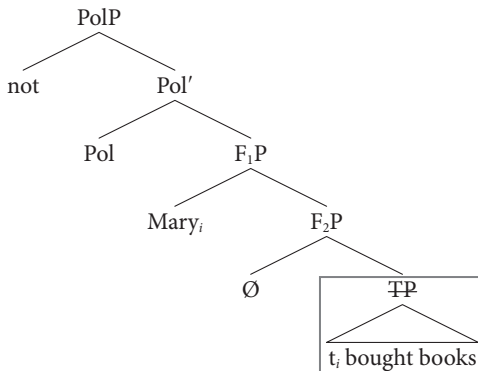
The *na Ayda* constituent adjoins to the DP *Araz*. However, since the DP *Araz* carries a contrastive focus feature, it moves to the right and adjoins to FocP.

Kolokonte (2008) proposed that Pseudo-stripping in English involves clausal coordination. Based on her analysis, the structure of the relevant parts of the English example in (39) would be as illustrated in (40).

(39) *John bought books, not Mary.*

(40) Structure of Pseudo-stripping in English

(Kolokonte 2008)



In (40), the DP *Mary* moves to the Spec of  $F_1P$  in the left periphery, before TP deletion takes place. In addition, the negative marker originates in the Spec of PolP. There are three main issues with Kolokonte's analysis. First, her proposal of two focus positions in the CP domain is not motivated.<sup>7</sup> Second, we know that information focus is not subject to overt movement to FocP (Kiss 1998).<sup>8</sup> Third, Kolokonte proposes that Pseudo-stripping involves a clausal coordination; however, since it is not possible to continue the sentence after *Mary*, as shown in (41), the TP undergoes an obligatory deletion.

(41) \*John bought books, not Mary bought books.

If Pseudo-stripping involves clausal coordination, we would expect to be able to coordinate it with another clause. However, this is not possible as the ungrammaticality of the sentence in (42) shows.

(42) \*John bought books, not Mary, neither Jack.

## 6. Conclusion

In this chapter, I have examined two Stripping constructions with negation: Polarity Stripping and Negative Stripping. I have shown that these constructions involve clausal coordination. Prior to ellipsis, a constituent that contrasts with its corresponding element in the preceding clause moves out of TP, which is specified for deletion. The remnant in Polarity Stripping and Negative Stripping moves to the

7. In her approach,  $F_1P$  hosts an element with a contrastive focus feature while  $F_2P$  hosts an element with an information focus feature.

8. She acknowledges that it is not possible for information focus to be preposed, as illustrated in (i).

- (i) Q: What did you order?  
 A1: I ordered pizza.  
 sA2: \*PIZZA I ordered. (Kolokonte 2008: 127)

However, following Baltazani's (1999) analysis of focus constructions in Greek, Kolokonte proposes that first the focused element moves to the left, and then the TP remnant moves to the left. Based on this proposal, the sentence in (iA1) is derived as illustrated in (ii).

- (ii) Step 1: Focus movement Pizza<sub>i</sub> [I ordered t<sub>i</sub>]  
 Step 2: TP-remnant movement [I ordered t<sub>i</sub> ]<sub>j</sub> pizza<sub>i</sub> t<sub>j</sub>

Step 2 renders the same word order as in (iA1). The focus movement is obscured by the subsequent movement of the remnant TP to a position higher than FocP, presumably TopP. Kolokonte bases her proposal on Greek data but she does not provide any English data. It is not clear whether focus fronting in Greek can be extended to English.

Spec of TopP and FocP, respectively. I propose that the negative marker in these constructions, which functions as a focusing adverb, originates in Pol head. In the second coordinate, the Pol head, which carries an [E] feature licenses the deletion of its complement, TP, at the PF level.

I also investigated another structure, Pseudo-stripping, that looks like Polarity Stripping and Negative Stripping. Based on the evidence from Persian, I argued that this structure does not involve ellipsis despite what has been claimed for English (Kolokonte 2008). I propose that Pseudo-stripping is derived via rightward movement and provide evidence that the negative marker in this structure is constituent negation and proposed that the NEG XP constituent originates adjacent to XP's corresponding element. Since it carries a contrastive focus feature, it moves to the right and adjoins to FocP, above *v*P, in the TP level.

The implications of these analyses are as follows: (a) the ellipsis site and the antecedent clause must have parallel information structure status, that is, both the remnant and its correlate must be either focalized or topicalized elements; (b) structures with *XP NEG* word order are bi-clausal while the structures with *NEG XP* word order are mono-clausal; (c) in structures with *XP NEG* word order, the remnant can be either a focalized or topicalized element while in structures with *NEG XP* word order, the remnant must be a focalized elements.

## Acknowledgements

Thanks to John Bailyn, Simin Karimi, Dan Finer, Thomas Graf, the audience at NACIL 1, and the reviewers for their feedback! I am also thankful to the Roshan Cultural Heritage Institute and especially its Chair and President, Dr. Elahé Omidyar Mir-Djalali for making this research possible.

## References

- Baltazani, Mary. 1999. Focus in Greek. Manuscript.
- Hankamer, Jorge & Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7. 391–426.
- Kahnemuyipour, Arsalan. 2001. On wh-questions in Persian. *Canadian Journal of Linguistics* 46. 41–61. <https://doi.org/10.1017/S000841310001793X>
- Karimi, Simin. 2005. *A minimalist approach to scrambling: Evidence from Persian*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110199796>
- Kiss, Katalin É. 1998. Identification versus information focus. *Language* 74. 245–273. <https://doi.org/10.1353/lan.1998.0211>
- Klima, Edward. 1964. Negation in English. In Fodor Jerry & Katz Jerrold (eds.), *The structure of language: Readings in the philosophy of language*, 246–323. Englewood Cliffs, NJ: Prentice Hall.

- Kolokonte, Marina. 2008. Bare argument ellipsis & information structure. Newcastle, UK: Newcastle University dissertation.
- López, Luis & Susanne Winkler. 2000. Focus and topic in VP-anaphora constructions. *Linguistics* 38. 623–664. <https://doi.org/10.1515/ling.2000.001>
- Kwak, Saera. 2010. Negation in Persian. *Iranian Studies* 43. 621–636. <https://doi.org/10.1080/00210862.2010.518028>
- Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.
- Rasekhi, Vahideh. 2018. Ellipsis and information structure: Evidence from Persian. Stony Brook, NY: Stony Brook University dissertation.
- Rasekhi, Vahideh. 2019a. Two types of *why* in syntax: Evidence from Why-stripping in Persian. Poster presented at *Sluicing and Ellipsis at 50*. Chicago, IL: University of Chicago.
- Rasekhi, Vahideh. 2019b. On the syntax of Why-stripping in Persian. Paper presented at *the 2nd North American Conference in Iranian Linguistics (NACIL2)*. Tucson, AZ: University of Arizona.
- Rasekhi, Vahideh. 2019c. Negative Stripping and Pseudo-stripping in Persian. Paper presented at *the 52nd Meeting of the Chicago Linguistic Society*. Chicago, IL: University of Chicago.
- Rizzi, Luigi. 1997. On the fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar, international handbook of linguistics*, 281–337. Kluwer: Dordrecht.
- Rooth, Mats. 1985. Association with focus. Amherst, MA: University of Massachusetts-Amherst dissertation.
- Rooth, Mats. 1996. *Focus*. In Shalom Lappin (ed.), *The handbook of contemporary semantic theory*, 271–297. Oxford: Blackwell Publishing.
- Taleghani, Azita. 2008. *The modality, aspect, and negation in Persian*. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.128>





# Oblique marking and adpositional constructions in Tat

## A mosaic of dialectal convergence and divergence

Murad Suleymanov

Université Paris Sciences et Lettres – École Pratique des Hautes Études /  
UMR 7192 “Proche-Orient – Caucase: langues, archéologie, cultures”

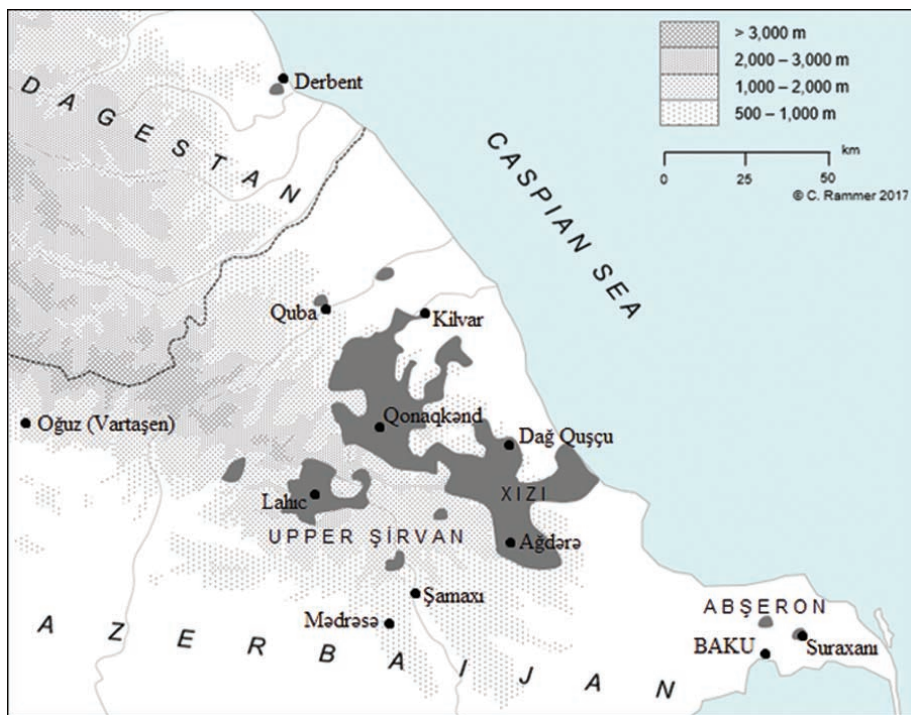
Tat, an Iranian language spoken in the Caucasus, has been exposed to heavy contact with neighbouring genetically unrelated languages, resulting in widespread bilingualism and causing contact-induced innovations. At the same time, a long period of low contact among the dialects of Tat has contributed to them displaying an abundance of phonological, lexical, morphological, and syntactic differences, as well as preserving otherwise extinct lexical elements and grammatical constructions. This article explores the various functions (inherited and introduced later on) of the Tat oblique clitic, with a focus on its role in forming new types of adpositional constructions.

**Keywords:** prepositions, postpositions, oblique marking, differential object marking, experiencer constructions, language contact, Azerbaijan, Tat, Juhuri, Azeri, Iranian, Armenian

### 1. Introduction

Tat, one of the lesser-known Iranian languages, is traditionally spoken in Azerbaijan and southern Russia (Dagestan) and is closely related to though not mutually intelligible with Persian. Due to a similarity in names, it is sometimes confused with the genetically much farther Tāti dialects of Iranian Azerbaijan. Today Tat is spoken in valleys along the slopes of the Greater Caucasus and all the way south to the Abşeron Peninsula, in a linguistically rich area where Indo-European languages meet East Caucasian and Turkic languages, resulting in extensive language contact.

Tat dialects are quite distinct from one another owing to limited contact among them. The latter is caused by their thin distribution in a relatively large area, coupled



Map 1. Distribution of Tat dialects (in dark grey)

with mountainous terrain (see Map 1), religious differences, migrations and growing language shift among its speakers, mainly to Azeri, a Turkic language and the official language of Azerbaijan. Tat dialects can be grouped into two main groups: Judeo-Tat and Muslim Tat.<sup>1</sup>

The aim of this article is to use oral and written Tat corpora to compare varieties of Tat from a specific morphological point of view. The focal point is the clitic  $=(r)A$ , referred to here as an oblique clitic. This clitic possesses a series of functions, giving rise to various grammatical structures. Section 2 will discuss oblique constructions in Tat, including background information necessary to understand their role in bringing about new adpositional constructions, addressed in § 3.

The functions of  $=(r)A$  discussed in this chapter include direct and indirect object marking, experiencer marking, and possessive marking in possessive constructions and possessive predicates, which in turn light the way for a description

1. These terms by no means reflect the religious affiliation of every speaker and are to be understood merely as denoting (sub-)linguistic grouping.

of its functions within adpositional phrases, including a special function termed ‘placeholder construction flagging’.

The dialects explored here include two Judeo-Tat varieties: literary (based on the dialect of Derbent, in Dagestan, southern Russia) and Vartaşen (spoken in northern Azerbaijan), and several Muslim Tat varieties: Şirvan (spoken in central Azerbaijan), Northern (spoken in northeastern Azerbaijan), Abşeron (spoken in eastern Azerbaijan), Xızı (spoken between the latter two), and Mədrəsə (previously spoken in central Azerbaijan).<sup>2</sup>

## 2. Oblique marking in Tat

Much has been said on the functions and distribution of the Persian morpheme *rā* (see Lazard 1982, 1994; Karimi 1990, 1996; among others, including most recently Jasbi, this volume and Karimi & Smith, this volume), which has been referred to as the subject of more specialised studies than any other Persian morpheme (Windfuhr 1979: 47). The existing literature includes diachronic treatments of *rā*, illustrating its morphosemantic development throughout the centuries (Lazard 1970; Paul 2008; Jügel 2019). The morpheme, whose origin dates back to Old Iranian (cf. Old Persian *rādi* “for the sake of”), has come to fulfil different functions in Western Iranian languages (Bossong 1985; Stilo 2004: 273; Paul 2017). In the case of the genetically closely related varieties of Tat, existing research does not account for the whole of the distinct and versatile values of this morpheme, often dialect-specific ones, including what appears to be recent developments.

---

2. Most of the data comes from oral and translated corpus collected between 2014 and 2017, as well as direct and indirect elicitations from native speakers. Examples containing a town/village name (e.g. “Məlhəm”) are extracts from that corpus. Other examples are quoted from published or unpublished scientific works: Authier (2012) for literary Judaeo-Tat, Grjunberg (1963) for Northern Tat, Lopatinskij (1894) and Ğalt’axc’yan (1970) for Mədrəsə Tat, Mammadova (2018) for Abşeron Tat, bearing in mind that the notation system and glosses might have been slightly modified for the purposes of uniformity of the data in this article. Examples extracted from non-scientific written corpora are cited as abbreviations of the author’s name and surname, the last two digits of the year of publication and a page reference (e.g. “MH09: 71”). Examples without references are either constructed or elicited for the purpose of this article.

The cognate of Persian *rā* in Tat is the oblique clitic  $=\langle r \rangle A$ ,<sup>3</sup> with the consonantal element disappearing if the clitic attaches to a consonant-final base.<sup>4</sup> Examples from Şirvan Tat: *xuna* “home” → *xuna=ra*, *şälä* “load” → *şälä=rä*, *kitab* “book” → *kitab=a*, *mäktäb* “school” → *mäktäb=ä*.

## 2.1 Differential object marking

Differential marking of both direct (DOM) and indirect (IOM) objects constitutes a prominent function of the clitic  $=\langle r \rangle A$ . This section will attempt to illustrate that IOM functions of this clitic are in some cases clearly inherited while in other cases seem to be the result of contact-induced secondary development of a lost inherited function.

### 2.1.1 Direct object marking

The value of  $=\langle r \rangle A$  as a clitic marking a direct object is parallel to that of modern Persian and is largely governed by referentiality. A definite direct object always carries the clitic:

#### (1) Şirvan Tat

a. *korda yof-t-um*.

knife find-PST-1<sup>5</sup>

“I found a knife/knives.”

b. *korda=ra yof-t-um*.

knife=OBL find-PST-1

“I found the knife (that I was looking for).”

An indefinite object may or may not be marked by  $=\langle r \rangle A$ . The marking depends on factors such as specificity and animacy, and the length of the verbal phrase following the object, which may trigger marking even on an indefinite or non-referenced object, as is also the case in Persian (Lazard 1994: 170).

3. The capital *A* indicates an alternation between *a* and *ä*. The sound transcribed as *a* corresponds to the vowel /a/ while *ä* corresponds to /æ/. Note that in some published corpora, the sound corresponding to /æ/ may be represented as *e* due to the sound /e/ being rare in some Tat varieties and generally being interpreted as an allophone of /æ/, or due to them both being interpreted as allophones of /ε/.

4. In some foothill Northern varieties, namely those of the villages of Afurca, Çiçi, Zərqava, Talabı, Qorxmazoba, Xaruşa, Səbətlər, İsnov, Zeyvə, Kilvar, Gəndov, Dağ Quşçu, Ərüsüküş, Püstəqasım, and Rustov, the clitic is  $=y\ddot{a}$  or  $=n\ddot{a}$  after vowels (except in pronouns and sometimes possessive clitics) and  $=\ddot{a}$  after consonants.

5. No number marking on the person implies the singular in all examples.

- (2) **Şirvan Tat** (Dəmirçi)  
*ye poyeyi=rä väyif-tän bā dās=i.*  
 one pole.IDF=OBL take-PRS:3 LOC hand=POSS:3  
 “He grabs a pole [with his hand].”
- (3) *bütün məslähät-o=ra bā-raf-tan-und<sup>6</sup> äz in käläʿämlä*  
 entire advice-PL=OBL IPFV-go-PRS-3PL from this PN  
*basto-ran-und.*  
 IPFV.get-PRS-3PL  
 “They go and get all advice from this Kalaʿamla.”

The clitic =(r)A as a direct object marker is characteristic of all Tat varieties.

### 2.1.2 Indirect object marking

Dative functions in Tat, including indirect object marking, are most commonly expressed by the dative-locative adposition (see § 3.1). However, some of these functions can also be fulfilled by the clitic =(r)A. This phenomenon is far from being universal in Tat varieties and is often semantically and syntactically conditioned.

In Judeo-Tat, the oblique marker=(r)A occurs with verbs of transfer and marks the indirect object when the direct object is indefinite and zero-marked (Authier 2012: 48):

- (4) **Judeo-Tat** (literary) (Authier 2012: 47)  
*zulpo mi-do väčä-ho=rä jüh vä sov.*  
 PN EVT-give.PST:3 chick-PL=OBL barley and water  
 “Zulpo would give the chicks barley and water.”
- (5) **Judeo-Tat** (Vartaşen)  
*äz ü ijaza xos-täd ki ä qurağ än vartaşin ye jigä*  
 from 3SG permission want-PRF.3PL SUB LOC edge of PN one place  
*d-ü uho=ra.*  
 MOD:give-3 they=OBL  
 “They asked his permission for him to give them land on the outskirts of Vartaşen.”

When the direct object is definite and takes the clitic =(r)A, the indirect object is marked by the locative adposition:

- (6) **Judeo-Tat** (literary) (Authier 2012: 186)  
*sovu=rä di ä rut.*  
 jug=OBL MOD:give:2 LOC PN  
 “Give the jug to Ruth.”

---

6. The grapheme *ä* corresponds to the central vowel /ə/.

However, when the indirect object is expressed by a pronoun, both objects may carry =(r)A (7) or only the direct object may do so (8) (Authier 2012: 186):

- (7) **Judeo-Tat** (literary) (HA82-T16: 34)  
*xär=tü=rä d-i mä=rä.*  
 donkey=POSS:2=OBL MOD:give-2 I=OBL  
 “Give me your donkey.”
- (8) **Judeo-Tat** (literary) (Authier 2012: 52)  
*mi-d-i ä mä sürx-ho=y=mä=rä.*  
 EVT-give-2 LOC I gold-PL=EZ=POSS:1=OBL  
 “You will give me my gold.”

In most Muslim Tat varieties, =(r)A-marked indirect objects are extremely rare. One notable case is the verb *birän* ‘to be’, which requires an oblique-marked indirect object when used in the sense of ‘happen to’ (cf. French *arriver* with the same semantics):

- (9) **Xızı Tat** (Ağdərə)  
*ti=rä çi bi-re?*  
 you=OBL what be-PREF.2/3  
 “What happened to you? / What came over you?” (elicited)

This construction exists in all Muslim Tat varieties, with the exception of Abşeron Tat.

The most widespread use of indirect object marking in =(r)A has been noted in the presently nearly extinct Tat variety of Mədrəsə, a village near Şamaxı and one of the two (along with Kilvar) villages populated before the Nagorno-Karabakh conflict by Tat speakers identifying as Armenian. Their dialect is classified as a variety of Muslim Tat (Miller 1929: 23), despite the historically Christian religious adherence of its speakers.

In Mədrəsə Tat, the dative use of =(r)A does not seem to be limited to verbs of transfer:

- (10) **Mədrəsə Tat** (Ğalt’axç’yan 1970: 85, 228, 271)  
*moy=i bā vaxt=i heyzad nā-vamuz-dā kilā=rä.*  
 mother=POSS:3 LOC time=POSS:3 nothing NEG-teach-PREF:3 girl=OBL  
 “At the time, her mother did not teach the girl anything.”
- (11) *u deşi bā uyekin otağ namak yazmiş kän-ä*  
 LOC enter.PST:3 LOC the.other room letter writing MOD:do-3  
*xuvär=štän=ä.*  
 sister=REFL=OBL  
 “He entered the other room to write a letter to his sister.”

In earlier works on Gilaki and Mazandarani, the use of the verb ‘to say’ with an addressee marked by a morphological element that generally marks a definite direct object led some researchers to describe said element as an accusative-dative marker (Rastorgueva & Ėdel’man 1982: 504; Šokri 1995 [1374]: 137; Shokri 2018). Recently, Ivanov & Dodyxudoeva (2017: 88) challenged this analysis, suggesting that the verb ‘to say’ and most other verbs from these languages described as having ‘dative-marked arguments’ simply treat the addressee/recipient as a direct object. According to them, the ‘dative reading’ only occurs when these verbs are translated into a language in which they are associated with dative semantics (e.g. the languages in which these descriptions are published). This view can likewise be applied to Mədrəsə Tat, which may very well have reanalysed the verbs in (48)–(49) as double-transitive, in contrast to all other Tat varieties.

In fact, the Mədrəsə Tat corpus used in this article was collected between 1953 and 1970 (Ġalt’axč’yan 1970: 9), at the time when Tat had been widely replaced by Eastern Armenian in Mədrəsə.<sup>7</sup> A strong Armenian identity, the existence of a school with Armenian as the language of instruction and intermarriage with Armenian speakers were among the factors that had contributed to growing bilingualism and eventually to massive language shift in the village.

In Eastern Armenian, the direct object of a transitive verb is marked by the dative case if it is human (Dum-Tragut 2009: 85):

- (12) **Eastern Armenian** (Dum-Tragut 2009: 86, 89)

*Ašot-ě tes-av Aram-i-n.*  
 PN-the see-AOR.3 PN-DAT-the  
 “Ashot saw Aram.”

- (13) *dasaxos-n usanoğ-ner-i-n bac’atr-um ē lezvabanut’y-an*

lecturer-the student-PL-DAT-the explain-PTCP:PRS COP:3 linguistics-DAT  
*nor tesakc’ut’yun-ner-ě.*  
 new theory-PL-the

“The lecturer explains new linguistic theories to the students.”

In addition, “verbs of transfer and verbs denoting mutual effect or a close connection with something/somebody” take an indirect object also in the dative case

---

7. Spreading knowledge of Armenian among the Tat speakers of Mədrəsə was mentioned as early as the end of the nineteenth century (Lopatinskij 1894: 30). In the 1920s, Tat had already been heavily marginalised as a language of everyday use, spoken by younger villagers among each other only when they did not want to be understood by Armenian speakers (Miller 1929: 13). In the 1950s, Mədrəsə Tat was described as nearing death (Ġalt’axč’yan 1957: 86). In 2014, a scholarly visit to Dprevank’, a village in Armenia where residents of Mədrəsə had relocated in 1989, did not discover any active speakers of Tat (Gilles Authier p.c. 2014).



(Dum-Tragut 2009: 87–89). The phrases in (10)–(11), therefore, when translated into Armenian (10')–(11'), will carry the dative marker:

- (10') **Eastern Armenian** (Ġalt'axč'yan 1970: 85, 229, 271)  
*mayr-ě žamanak-i-n aġč'ka-n voč'inč' č'i*  
 mother-the time-DAT-the girl.DAT-the nothing NEG.COP:3  
*sovore-cr-el.*  
 teach-CAUS-PTCP:PRF  
 “At the time, the mother did not teach the girl anything.”
- (11') *na mt-av myus senyak vor namak gr-i ir*  
 3SG enter-AOR.3 other room SUB letter write.SBJV-3 own  
*k'uyr-ik-i-n.*  
 sister-DIM-DAT-the  
 “He entered the other room to write a letter to his sister.”

Since in the two late nineteenth-century Mədrəsə Tat texts published in Lopatinskij (1894: 25–30), no cases of  $=(r)A$  for dative functions are found,<sup>8</sup> it is safe to assume that indirect object marking in  $=(r)A$  in this dialect is an innovation rather than an inherited function. It is likely that the Armenian adstratum resulted in the speakers' likening the Tat oblique marker  $=(r)A$  to the Armenian direct-object-marking dative and thus extending the function of  $=(r)A$  in accordance with the distribution of the dative marking in Armenian (functional extension rather than replacement is evident from the fact that Mədrəsə Tat has not relinquished the original 'non-Armenian' functions of  $=(r)A$ , such as non-human direct object marking). Further examples show that the development of ditransitives with both arguments marked by  $=(r)A$  had not been completed in Mədrəsə Tat as of 1953, since the same verb used in similar semantic contexts demonstrates varying valency ((14) vs. (15)), sometimes within the same utterance (16):

- (14) **Mədrəsə Tat** (Ġalt'axč'yan 1970: 229, 89, 270)  
*horomsim šüvār=štän=ä mǎ-fürsä-rän moy-piyār=štän=ä.*  
 PN husband=REFL=OBL IPFV-send-PRS:3 mother-father=REFL=OBL  
 “Horomsim sends her husband over to his parents' (house).”
- (15) *mǎn ye kaġat fürs-üm bǎ hambaz=mǎn.*  
 I one paper MOD:send-1 LOC friend=POSS:1  
 “Let me send a note to my friend.”

---

8. One of those texts was checked with a native speaker in 1928, and her version was published in Miller (1945: 124–125). This latter version did not contain instances of the indirect object marker  $=(r)A$  either, though the speaker may have been influenced by the older text.

- (16) *ki be män ye araba giyov t-ä män u=ra ye xäm*  
 who LOC I one cart grass MOD:give-3 I 3SG=OBL one jug  
*šäräb m-t-äm.*  
 wine EVT-give-1  
 “To him who gives me a cartful of grass, I will give a pitcherful of wine.”

In their critical assessment of recent hypotheses that tend to ascribe change within a language spoken in a bi- or multilingual area to language contact instead of considering language-internal development, Poplack & Levey (2010: 398) offer particular criteria to safely assume contact-induced change:

1. that the change be absent in the pre-contact and non-contact variety;
2. that the change, if present in the pre-contact and non-contact variety, not be conditioned in the same way as in the presumed source variety *and*
3. that the change parallel the behaviour of a counterpart feature in the source variety.

The first criterion appears to be rather ambiguous for Mədrəsə Tat. On the one hand, indirect object marking in  $=(r)A$  is absent in texts recorded at the time when contact with Eastern Armenian was weaker. On the other hand, this dialect, like all of Tat, almost certainly descends from an Iranian variety that was familiar with marking indirect objects with a cognate of  $=(r)A$ , as illustrated by this feature still persisting in Judeo-Tat. Nevertheless, a historical feature resurfacing and gaining prominence after a long period of absence does not look like a convincing scenario.

To back this up, the second and third criteria should be taken into account. While in Judeo-Tat, indirect object marking in  $=(r)A$  is limited to specific syntactic conditions, its distribution in Mədrəsə Tat seems to be more haphazard. Where it is encountered, it patterns argument marking of Eastern Armenian, and this type of indirect object marking appears to be more common than in any other Tat variety that features it. This together with Mədrəsə Tat’s relative isolation from other Tat varieties and its distinct sociolinguistic status suggests that the twentieth-century use of  $=(r)A$  as a common indirect object marker can be considered a contact-induced development with a reasonable degree of certainty.

## 2.2 Experiencer function

The use of  $=(r)A$  to mark the experiencer in constructions involving affective predicates, with the stimulus in the nominative, is found across different varieties of Tat, but is more typical of Judeo-Tat. The robust use of this inherited construction may be bolstered by a heavier contact between Judeo-Tat and East Caucasian languages,

where such constructions are common, while Muslim Tat varieties are mainly exposed to Azeri, where affective predication is characterised by a nominative-subject construction or an adverbial construction with the stimulus in the nominative.

- (17) **Judeo-Tat** (literary) (Authier 2012: 212–213)  
*tü=rä zä-rä xäyol=i mä=rä.*  
 you=OBL hit-INF dream=COP:2/3 I=OBL  
 “To me, killing you is a dream.”
- (18) *padšoh=ä äz šori jığä nist=i.*  
 king=OBL from joy place NEG.COP=COP:2/3  
 “The king could not be happier.”
- (19) *diyä u=rä bäq gäräk nist=i.*  
 more 3SG=OBL frog necessary NEG.COP=COP:2/3  
 “He does not need the frog anymore.” [*lit.* “The frog is not necessary for him anymore.”]

In Muslim Tat, such marking is found mainly in predicates expressing physical affection, such as experiencing hunger or thirst, or a reaction to outside temperature:

- (20) **Northern Tat** (Qonaqkənd) (MH09: 71)  
*amu xištän nä-bas-tanbü gu-nü ki ü=rä*  
 uncle self NEG-IPFV.know-IMPF.3 SJBV:say-3 SUB 3SG=OBL  
*kisnä bi-re.*  
 hungry be-PRE.3  
 “His uncle could not say that he had become hungry.”
- (21) **Abşeron Tat** (Suraxanı) (Mammadova 2018: 28)  
*därveçä=rä best, xirik=ü ayol=ä.*  
 window=OBL MOD:close:2 cold=COP:3 child=OBL  
 “Close the window, the child is cold.”
- (22) **Şirvan Tat** (Əhən)  
*unja be män yə kandiçaner=i bil-ind, män=ä gər=i.*  
 there for I one air\_conditioner=IDF MOD.put-2PL I=OBL hot=COP:2/3  
 “In that case, install an air conditioner for me (because) I am hot.”

There is a limited set of modal expressions in Muslim Tat which license  $=(r)A$ -marked experiencers, notably the verbal compounds *gäräy(=i) birän* and *lozim birän* “to be required” (though the latter two are more frequently used with a dative-locative preposition):

(23) **Şirvan Tat** (Gombori)

*ägär işmun=a gārây=i hi äz in şüş*  
 if you.PL=OBL necessity=POSS:3 EXIST.COP:2/3 from this trellis  
*bär-ind.*

MOD:carry-2PL

“If you need (beans) [*lit.* “if to you, there is a need of beans”], take (some) from this trellis.”

(24) **Şirvan Tat** (Qoydan)

*umun=a çi lozim=i?*

we=OBL what required=COP:2/3

“What do we need?”

Special mention should be made regarding the verb *vöüstän* “to be wished for”, likely a denominative verb whose root is related to Persian *viār* “longing, craving”. In the varieties where it exists (both Jewish and Muslim), this verb requires an oblique-marked experiencer:

(25) **Şirvan Tat** (Burovdal)

*näne=män=ä xämzä vö-üst.*

grandmother.EZ=POSS:1=OBL melon be\_wished-PST:3

“My grandmother craved melons.”

(elicited)

### 2.3 Possessive constructions

Tat varieties generally employ more than one possessive construction, though not necessarily the same set. A construction encountered across all the varieties – a cognate of the Persian *Ezafe* construction – is the one where the possessor is post-posed to the possessed. Note that Tat has lost the *Ezafe* marker but traces of it can still be seen on *ä-* or *a-*final head nouns (in Judeo-Tat, on all vowel-final nouns) in the form of a word-final fronted element (26b), (27b), glossed as EZ (for ‘*Ezafe*’) for etymological reasons.

(26) **Judeo-Tat**

a. *biror saro*

brother PN

“Sara’s brother”

b. *xuney saro* cf. *xunä*

home.EZ PN home

“Sara’s house”

(27) **Şirvan Tat**

- a. *birör sora*  
 brother PN  
 “Sara’s brother”
- b. *xune sora* cf. *xuna*  
 home.EZ PN home  
 “Sara’s house”

The possessor can be encoded by a personal (possessive) clitic on the possessed. The possessive clitics of Tat differ in form across its varieties but have mostly evolved from personal pronouns attached to their heads as they would in an *Ezafe* construction in (26)–(27).

- (28) a. **Northern Tat (Kilvar)** (Ġalt’axc’yan 1970: 254)  
*bürar=mä* cf. *män*  
 brother=POSS:1 I  
 “my brother”
- b. **Şirvan Tat**  
*xune=şmun* cf. *işmun*  
 home.EZ=POSS:2PL you.PL  
 “your house”

In some varieties, constructions such as those in (28) can additionally involve an oblique-marked possessor while maintaining the possessive clitic on the possessed (*lit.* ‘X’s Y’ = ‘of X his/her/its Y’). This possessor can be both nominal (29a) and pronominal (29b):

(29) **Şirvan Tat**

- a. *sora=ra birör=i*  
 PN=OBL brother=POSS:3  
 “Sara’s brother”
- b. *işmun=a xune=şmun*  
 you.PL=OBL home.EZ=POSS:2PL  
 “your house”

It is without a doubt that such double-marked possessive constructions result from centuries-long contact between Tat and Azeri, where possessives are likewise double-marked:

(29’) **Azeri**

- a. *Sara-nın qardaş-ı*  
 PN-GEN brother-POSS:3  
 “Sara’s brother”

- b. *siz-in ev-iniz*  
 YOU.PL-GEN home-POSS:2PL  
 “your house”

A Tat possessive construction like the one in (30) will hereafter be referred to as ‘oblique-marked’. Note that if an oblique-marked possessive construction involves an adposition, the latter is placed between the possessor and the possessed (30a)–(30b):

(30) **Şirvan Tat**

- a. *sora=ra ä birör=i*  
 PN=OBL from brother=POSS:3  
 “from Sara’s brother”
- b. *işmun=a ba xune=şmun*  
 YOU.PL=OBL LOC home.EZ=POSS:2PL  
 “to/in your house”

Oblique marking of this type is also used to lexicalise two-noun phrases, in which the dependent has a property-like function. Unlike possessive constructions, this seems to be an internal development of Tat (in Azeri, in this case, there would be no marking on the ‘property’).

(31) **Şirvan Tat**

- a. *şel=ä boğče=yi*  
 child=OBL garden.EZ=POSS:3  
 “nursery, kindergarten”
- b. *xuna=ra ħēvun=i*  
 home=OBL animal=POSS:3  
 “domestic animal”

Oblique-marked possessives such as those in (30) are found in all Muslim Tat varieties except Abşeron Tat. They are absent in literary Judeo-Tat, but are notably present in Vartaşen Judeo-Tat:

(32) **Judeo-Tat (Vartaşen)**

- zeynāb\_xanlarova boku=ra äz doğli\_mähälle=yi=ni.*  
 PN PN=OBL from PN\_quarter.EZ=POSS:3=COP:2/3  
 “Zeynāb Xanlarova is from the Dağlı quarter of Baku.”

- (33) *vartaşin=a jūhur-o=yi=ra zuhun=i uho=roz*  
 PN=OBL Jew-PL=POSS:3=OBL language=POSS:3 they=with  
*toşni yeki=ni.*  
 exactly one=COP:2/3

“The language of the Jews of Vartaşen is exactly the same as theirs.”

In (33), one can identify two oblique-marked possessive constructions: *vartašin=a ĵuhur-o=yi* “the Jews of Vartašen” and *ĵuhur-o=ra zuhun=i* “the language of the Jews”. This can schematically be represented as:  $[[vartašin=a \text{ } \acute{z}uhur-o=yi]=ra \text{ } zuhun=i]$ , where the second  $=(r)A$  marks a possessor expressed by a possessive phrase.

Among Iranian languages, a similar possessive construction has developed in Northern Tajik, also in contact with a Turkic language (Uzbek). Here, it widely replaces the Persian *Ezafe* construction:

- (34) **Northern Tajik** (Windfuhr & Perry 2009: 443)
- a. *man=a pisar=am*  
 child=RA son=POSS:1  
 “my son”
- b. *Zaydullo=ra palink-o=š*  
 PN=RA shoe-PL=POSS:3  
 “Zaydullo’s shoes”

In general, the choice between an *Ezafe* possessive construction and an oblique-marked possessive construction in Tat is pragmatic. According to one description, an *Ezafe* possessive construction is more marked in use and typical of constructions characterised by a “close natural connection between the constituents” (Grjunberg 1963: 25). More specific reasons behind this choice need to be investigated using a detailed corpus analysis, which falls outside of the scope of this study.

## 2.4 Possessive predication

Similar to possessive constructions, possessive predicates manifest themselves differently across Tat varieties.<sup>9</sup>

All Tat varieties have some form of oblique-marked existential possessive predication. In this type of predication, the possessor is marked by  $=(r)A$  and the possessee acts like the morphological subject, which agrees with the copula. In all, the construction, which can be literally translated as ‘to POSSESSOR, there is POSSESSEE’, is reminiscent of the Latin *mihi est*-type predication and is inherited in Tat.

- (35) **Abşeron Tat** (Suraxanı) (Mammadova 2018: 82)
- padšah=ä yä-to-lä kuk bi-rän.*  
 king=OBL one-QTF-DIM son be-PRS:3  
 “The king had only one son.” (narrative PRS)

9. This section will not focus on transitive HAVE-like predicates that have developed in some varieties (Abşeron Tat, Mədrəsə Tat) and are used alongside existential predicates.

- (36) **Judeo-Tat** (literary) (Authier 2012: 148)  
*mä=rä sä kuk=i.*  
 I=OBL three son=COP:2/3  
 “I have three sons.”

Nevertheless, the two varieties above are a minority in this respect. In Şirvan Tat and Northern Tat, a similar construction has only been encountered in exclamative sentences where the function of  $=(r)A$  is close to that of the benefactive:

- (37) **Şirvan Tat** (Burovdal)  
*çi xub ki in kitob män=ä hist=i!*  
 what good SUB this book I=OBL EXIST=COP:2/3  
 “It is so good that I have this book!” (elicited)

In most Tat varieties, oblique-marked possessive predicates require a possessive clitic on the possessee:

- (38) **Şirvan Tat** (Əhən)  
*umun=a yēloğ=mun hist=i.*  
 we=OBL summer\_pasture=POSS:1PL EXIST=COP:2/3  
 “We have a summer pasture.”
- (39) **Mədrəsə Tat** (Lopatinskij 1894: 25)  
*takawor=a bi-rä sä kük=i.*  
 king=OBL be-PRF:3 three son=POSS:3  
 “The king had three sons.” (narrative PRF)
- (40) **Judeo-Tat** (Vartaşen)  
*mä=rä häft div gedäy=mä hist=i.*  
 I=OBL seven ogre boy.EZ=POSS:1 EXIST=COP:2/3  
 “I have seven ogre sons.” [KN]

This double-marked possessive predication is very similar to the Azeri one, which is likely responsible for its development in Tat.

- (38') **Azeri**  
*bizim yaylağ=ımız var.*  
 we.GEN summer\_pasture=POSS:1PL EXIST:3  
 “We have a summer pasture.”

As in Azeri, when the oblique-marked possessor is not topical, it can be dropped without bringing about ambiguity because the possessee maintains the co-referential possessive clitic:



- (38'') **Şirvan Tat**  
*yēloğ=mun* *hist=i.*  
 summer\_pasture=POSS:1PL EXIST=COP:2/3  
**Azeri**  
*yaylağ=ımız* *var.*  
 summer\_pasture=POSS:1PL EXIST:3  
 “We have a summer pasture.”

It should be noted that in the Northern Tat variety of Dağ Quşçu and in Xızı Tat, one comes across both zero-marked and clitic-marked possessives.

- (41) Northern Tat (Dağ Quşçu) (Grjunberg 1963: 35, 73)
- a. *hāmin* *čuban=ä* *ye* *kuk-lä* *im-bi-rän.*  
 this\_very shepherd=OBL one son-DIM IPFV-be-PRS:3  
 “This shepherd had a son.” (narrative PRS)
- b. *mān=ä* *yetakä* *kuk=mān* *hist.*  
 I=OBL only\_one son=POSS:1 EXIST:3  
 “I have only one son.”

Grjunberg, who conducted his research in the 1950s, does not suggest an environment that would warrant one or the other construction. Ağacamal Soltanov (p.c. 2018), one of the authors of Soltanov & Soltanov (2013) and a native speaker of the variety in (41), characterises zero-marked possessives, such as that in (41a), as lower in referentiality. According to him, they are rather infrequent today and their uncommonness should be ascribed to growing Azeri influence.

Northern Tat of Dağ Quşçu and Xızı Tat are the only varieties where two competing oblique-marked predication constructions have been found so far. Consulted speakers of Şirvan Tat and of other Northern varieties do not accept zero-marked possessives as grammatical outside of sentences like (37).

### 3. Adpositional constructions in Tat

An overview of adpositions is required in order to pave the way for a more comprehensive presentation of oblique constructions in Tat. This section focuses in particular on the Muslim Tat of Upper Şirvan, spoken in north-central Azerbaijan, in an area directly to the south of the Greater Caucasus mountain ridge (see Map). Adpositions in Şirvan Tat can be divided into two main types: simple and compound.

### 3.1 Simple adpositions

Şirvan Tat has the following simple adpositions:

**Table 1.** Simple adpositions of Şirvan Tat

Adposition type	Şirvan Tat
Dative-locative “in, to”	<i>bä</i> (with phonetically conditioned variants <i>bä, be, ba</i> )
Ablative “from”	<i>ä</i> <i>äz</i> (before pronouns and demonstratives)
Comitative-instrumental “with, by, through, along”	<i>vo, ve</i> (used interchangeably)

The use of the first two adpositions is illustrated in (42)–(43):

(42) **Şirvan Tat**

*bä häyot ye kârg=i hi.*

LOC yard one chicken=IDF EXIST:2/3

“There is a chicken in the yard.”

(43) *fürmo-re ä äs.*

descend-PRF.2/3 from horse

“He dismounted from his horse.”

Note that when combined with personal pronouns, simple adpositions can fuse with the latter, thus *bä* “to” + *ü* “s/he, it” results in *bö* “to him/her/it” and *vo/ve* “with” + *ü* “s/he, it” results in *vö* “with him/her/it”.

### 3.2 Compound adpositions

Compound adpositions are adpositional phrases grammaticalised to various extents, which express more specific spatial relations than simple adpositions. They consist of two elements: a simple adposition (one of those described in Table 1) and a noun indicating a place. Below is an exemplary list of compound adpositions in Şirvan Tat:

**Table 2.** Some compound adpositions of Şirvan Tat

---

<i>ba ara</i>	“between, among” ( <i>ara</i> “gap”)
<i>bä birun</i>	“outside of” ( <i>birun</i> “outdoors”)
<i>bä darun</i>	“inside of” ( <i>darun</i> “indoors”)
<i>bä kinor</i>	“near, beside” ( <i>kinor</i> “edge”)
<i>bä mingäh</i>	“in the middle of” ( <i>mingäh</i> “middle”)
<i>bä pähli</i>	“next to” (* <i>pähli</i> “side”)
<i>bä pišö</i>	“in front of” ( <i>pišö</i> “front”)
<i>bä pišt</i>	“behind” ( <i>pišt</i> “back”)
<i>bä sär</i>	“on top of” ( <i>sär</i> “head”)
<i>bä tän</i>	“toward, at, next to” ( <i>tän</i> “body”)
<i>bä yon</i>	“near, beside” ( <i>yon</i> “side”)
<i>bä zir</i>	“under” ( <i>zir</i> “bottom”)
<i>vo/ve darun</i>	“through” ( <i>darun</i> “indoors”)
<i>vo/ve riz</i>	“after, following” (* <i>riz</i> “sequence?” <sup>10</sup> )

---

Compound adpositions are an open class. The combination ‘adposition + noun’ is productive and can generate a large number of new compounds, such as *ä darun* ‘from within’, *vo sär* ‘along the top of’, etc. These can be used as adpositions when used with a nominal head (44a) or as adverbs when used without a nominal head (44b):

## (44) Şirvan Tat

- a. *raf-t-um bä\_zir kirpi.*  
 go-PST-1 under bridge  
 “I went under the bridge.”
- b. *raf-t-um bä\_zir.*  
 go-PST-1 under  
 “I went below/underneath.”

A compound adposition consists in part of a nominal element (hereafter ‘internal complement’) and it combines with its nominal head (hereafter ‘external complement’) in the form of a possessive construction. As with possessives, there is an *Ezafe* construction (prepositional) vs. an oblique-marked (postpositional) construction dichotomy in the case of compound adpositions. The distribution of these two constructions in the Şirvan Tat corpus is roughly equal to the distribution of the two possessive constructions in (28) and (30).

---

10. A parallel between \**riz* and Sorani *riz* ‘sequence’ has been suggested by Thomas Jügel (p.c. 2017).

### 3.2.1 *Ezafe compound prepositional constructions*

In an *Ezafe* compound prepositional construction, the compound adposition is preposed to its external complement as in the possessive construction in (28). The construction suggests that this form is treated as a single adposition by the speakers in terms of the syntax. Note that in (49), an *a*-final head noun undergoes the same vowel change as the head in (28b):

- (45) **Şirvan Tat** (Dəmirçi)  
*bä\_darun qäzit=iş ye jif qaloş poy vor-de.*  
 inside newspaper=ADD one pair clog share bring-PRF.2/3  
 “And inside the newspaper, he brought a pair of clogs as a present.”
- (46) *bä\_sär qäbr-ho nöüş-tond.*  
 on tomb-PL write-PRF.3PL  
 “It is written on the tombstones.”
- (47) [“May God always send food to your door...”]  
*bä dür nä-dö-yi vo\_ruz nun.*  
 LOC far NEG-MOD:run-2 following bread  
 “...so that you do not have to run far in search of bread!”
- (48) [“They invest him in so many robes that...”]  
*ba-mun-dan bä\_zir xälät.*  
 IPFV-remain-PRS:3 under robe  
 “...he ends up under (a pile of) robes.”
- (49) *in sutun ba-mun-dan ba\_are biläg-ho=yi.*  
 this column IPFV-remain-PRS:3 between.EZ arm-PL=POSS:3  
 “The column ends up between her arms.”

### 3.2.2 *Oblique-marked compound postpositional constructions*

In an oblique-marked compound postpositional construction, the external complement (nominal or pronominal), when it is overt, receives the clitic =(r)A while the nominal element of the adposition is marked by a possessive clitic. Examples (50)–(54) show the behaviour of the same prepositions as in (45)–(49) in oblique-marked constructions:

- (50) **Şirvan Tat** (Burovdal; Dəmirçi)  
*ye däqqä=rä bä\_darun=i här kor däğiş bi.*  
 one minute=OBL inside=POSS:3 each work changing be.PST.3  
 “Within one minute, everything changed.”
- (51) *tojir lüt\_şüryun taxa=rä bä\_sär=i xisi-re.*  
 merchant naked.STR board=OBL on=POSS:3 sleep-PRF.2/3  
 “The merchant, naked as a jaybird, is sleeping on the board.”

- (52) *yek=i=rä ve\_ruz=i hazor-to biye!*  
 one=POSS:3=OBL following=POSS:3 thousand-QTF MOD.come.2/3  
 “Following one of them, may a thousand come!”
- (53) *qurbun bu-bur-um ü=rä bä\_zir=i.*  
 sacrifice MOD-cut-1 3SG=OBL under=POSS:3  
 “...that I slaughter a sacrificial lamb under it (i.e. the body).”
- (54) *ve rusyät umun=a ba\_are=mun dih xëbäri=yi.*  
 with PN we=OBL between.EZ=POSS:1PL village PN=COP:2/3  
 “(Located) between Russia and us is the village of Xeybäri.”

### 3.3 Placeholder construction

The existence of oblique-marked compound postpositional constructions may have contributed to the emergence of a placeholder construction in Şirvan Tat. ‘Placeholder construction’ is a term applied by Jügel (2016) to a phenomenon observed in Middle Persian as illustrated in (55):

- (55) **Middle Persian** (Jügel 2016: 50)  
*u=š dām dō ēk merd ud ēk zan az=iš*  
 and=PC:3 creature two one man and one woman from=EXPL  
*dēs-ād.*  
 build-2PL.SBJV  
 “... and shall form two creatures – a man and a woman – out of it”

In a placeholder construction,

the enclitic pronoun appears in Wackernagel position and the preposition that controls it is followed by an expletive [third-person singular enclitic pronoun]. This means that there are two enclitic pronouns. The [third-person singular enclitic pronoun] that is attached to the preposition does not refer to an actant. Instead, it takes the position of the control[l]ed [enclitic pronoun] that seems to be fronted, i.e. it holds the [place of the enclitic pronoun] after the preposition.

(Jügel 2016: 50)<sup>11</sup>

A peculiar construction involving  $=(r)A$  has developed in one specific Tat variety, that of Şirvan Tat. In some of its sub-varieties, a special adpositional construction has been noted: the pronoun *ü* ‘s/he, it’ attaches to a simple adposition (recall that *bö* and *vö* are contracted forms of *bä* and *vo/ve* with the third-person singular pronoun) which follows an oblique-marked complement:

11. The term ‘Wackernagel position’ refers to the position after the first word or phrase in a clause.

- (56) Şirvan Tat (Lahıc, Həftəsov, Əhən, Burovdal)  
*hämum=a äz=ü dü-to min-de.*  
 bathhouse=OBL from=3SG two-QTF stay-PRF.2/3  
 “Two (of the) bathhouses are left.”
- (57) *minjivon=a bö pö-üst-und.*  
 PN=OBL LOC.3SG stand-PST-3PL  
 “They stopped at Mincivan.”
- (58) *bä çäki=ti bi-rä üdömin-ho=ra vö bi-niş*  
 LOC weight=POSS:2 be-PTCP person-PL=OBL with.3SG MOD-sit:2  
*vaxs-i!*  
 MOD\_get.up-2  
 “Associate [*lit.* sit and get up] with people of your own level [*lit.* weight]!”
- (59) *jürbäjür num-ho=yi bi-rä tärsi-räni-yä çi-ho=ra*  
 different name-PL=POSS:3 be-PTCP fear-GRDV-ATR thing-PL=OBL  
*üz=ü ixtilot bə-sox-tanbi.*  
 from=3SG narration IPFV-do-IMP.F.3  
 “He would tell lots of stories about scary creatures of different names.”

From a morphological point of view, it is important to underline that the constructions in (56)–(59) are different from the oblique-marked compound postpositional constructions in (50)–(54), which are also characterised by double marking. First, the constructions in (56)–(59) are restricted to cases where the complement refers to the third person (as opposed to (54), for instance). Second, the post-adpositional reference to the complement (i.e. the placeholder) can be expressed only by the personal pronoun *ü* and not by possessive clitics:

- (56') *hämum=a äz=ü (\*üz=i) dü-to min-de.*  
 bathhouse=OBL from=3SG ( from=POSS:3) two-QTF stay-PRF.2/3  
 “Two (of the) bathhouses are left.”

When comparing (56)–(59) with the Middle Persian example in (55), one can notice a morphological parallel: a semantically redundant third-person enclitic pronoun (placeholder) attaching to (what is originally) a preposition and found in the morphological slot of an argument expressed earlier in the sentence. At the same time, there are noticeable differences. In Middle Persian, the placeholder element *=iš* can be co-referent with a complement not only in the third person (Jügel 2016: 50). Furthermore, the Middle Persian complement can be topicalised (Jügel 2016: 52–53) and dislocated very far from the preposition, whereas in Tat, there are no cases of lexical or morphological elements being inserted between the oblique-marked complement and the adposition; the construction is quite fixed despite its likely recent origin (see below).

The existence of constructions in (56)–(59) can hardly justify referring to Tat simple adpositions as “alternating adpositions”, defined as “identical words that function either as prepositions or postpositions depending on their meaning or function” (Stilo 2005: 52)<sup>12</sup> because the latter condition is not fulfilled. Elicitation and corpus analysis have indicated that the speakers did not perceive semantic differences between the constructions with preposed and postposed simple adpositions. In addition, postposed simple adpositions cannot occur without the oblique marker on the noun, nor can they be preposed to an oblique-marked noun.

It is important to stress that the comparison of Tat with Middle Persian in no way suggests a historical link between the two phenomena. In fact, cross-dialectal comparison indicates that placeholder constructions are a recent phenomenon in Tat. Not only are they absent outside of Şirvan, but even within this region, they are not universally accepted (although where accepted, they are rather frequently used). Two arguments could help date the emergence of the placeholder construction in Tat to the twentieth century: its absence in Gombori, a Tat-speaking village in present-day Georgia settled by immigrants from Lahıc and Əhən (two locales where placeholder constructions are found today) a century ago, as well as its absence in the early-twentieth-century corpora from Lahıc.

Heavy contact with Azeri, a language with no prepositions but a rich set of postpositions and case suffixes, may have been key in causing Tat to develop placeholder constructions. If so, this was facilitated by the fact that Tat already had at its disposal constructions where adpositional constructions were postposed to their complements (see 3.2.2). Simple adpositions could not be combined with possessive clitics, hence a more typical ‘simple adposition + personal pronoun’ formula was chosen instead. In this regard,  $=(r)A$  receives a new interpretation: that of a flag signaling a right-branched simple adposition.

A historically prepositional language where prepositions have come to display parallel uses as both prepositions and postpositions due to language contact is not a rare occurrence. Among Iranian languages, this has been noted for Balochi dialects, with the most extreme case being the dialect of Karachi, Pakistan, which has become postpositional due to heavy contact with likewise postpositional Indo-Aryan languages (Farrell 2003: 196). In the case of Tat, it is thus the mobility of the historical prepositions that has provoked the use of the looser term ‘adposition’ in this article. The phenomenon in Tat is typologically remarkable in that it reflects a syntactic permutation rendering an inherited syntactic structure heavier.

---

12. Such elements are found in some Tati dialects, in Zazaki, Gorani, as well as in some Central Iranian languages (Stilo 2005: 52)

## 4. Conclusion

Table 3 is a recapitulative table generated based on the functions of the oblique marker  $=\langle r \rangle A$  in the analysed Tat varieties. ‘(+)’ indicates that the function exists but is limited to very specific cases.

**Table 3.** Functions of  $=\langle r \rangle A$  in Tat varieties

Function	Variety					
	Judeo-Tat		Muslim Tat			
	Literary	Vartaşen	Mədrəsə	Şirvan	Northern	Abşeron
Direct object marker	+	+	+	+	+	+
Indirect object marker	+	+	+	(+)	(+)	–
Experiencer marker	+	+	+	+	+	+
Possessor marker (incl. internal complement of an adpositional phrase)	–	+	+	+	+	–
Possessor marker in a possessive predicate (possessee unmarked)	+	–	–	(+)	(+)	+
Possessor marker in a possessive predicate (possessee marked)	–	+	+	+	+	–
Placeholder construction flag	–	–	–	+	–	–

The data in Table 3 shows that direct object marking and experiencer marking are the sole functions of  $=\langle r \rangle A$  that are common to all the varieties. It should be stressed that although parallels are observed for some other functions, development paths may not necessarily be identical. Such is the case of the indirect object marker function of  $=\langle r \rangle A$  whose existence in Judeo-Tat and to a limited extent in most Muslim Tat is historically unrelated to a parallel function in Mədrəsə Tat. The table also shows Şirvan Tat standing out as the only variety that uses  $=\langle r \rangle A$  in placeholder constructions and in general, displaying the most versatile use of the oblique marker.

## Acknowledgements

I would like to thank Gilles Authier, Sebastian Fedden, Thomas Jügel, Agnes Korn, as well as all my anonymous reviewers for their valuable comments and suggestions, which helped me improve this article.



## List of abbreviations

ADD	additive	NEG	negative
AOR	aorist	OBL	oblique
ATR	attributive	PC	personal clitic
COP	copula	PL	plural
DAT	dative	PN	proper noun
DIM	diminutive	POSS	possessive
EVT	eventual	PQP	pluperfect
EXIST	existential marker	PRF	perfect
EXPL	expletive pronoun	PST	perfective past
EZ	<i>Ezafe</i>	PTCP	participle
GEN	genitive	QTF	quantifier
IDF	identification marker	RA	Persian <i>rā</i>
IMPF	imperfect	REFL	reflexive
INF	infinitive	SBJV	subjunctive
IPFV	imperfective	STR	strong adjective
LOC	locative	SUB	subordinator
MOD	modal		

## Tat corpus bibliography

- [HA82–T16] Avšalumov, Hizgil (ed.). 1982. *Xitroumnyj Šimi Derbendi* [Crafty Šimi of Derbend]. Makhachkala: Dagestankoe knižnoe izdatel'stvo (quoted in *Tovuši* #74, February 2016, p. 34)
- [KN] *Keçəlin nağılı* [A Bald Man's Tale]. Translated into various dialects.
- [MH09] Hacıyev, Maqsud. 2009. *Azərbaycan tatlarının dili* [The Language of the Azerbaijani Tats]. Baku: Mütərcim.

## References

- Authier, Gilles. 2012. *Grammaire juhuri, ou judéo-tat, langue iranienne des Juifs du Caucase de l'est* [A grammar of Juhuri or Judeo-Tat, the Iranian language of the Jews of the East Caucasus]. *Beiträge zur Iranistik* 36 [Bibliothèque iranienne 76]. Wiesbaden: Reichert.
- Bossong, Georg. 1985. *Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen* [Empirical Universals Research: Differential Object Marking in New Iranian Languages]. Tübingen: Narr.
- Dum-Tragut, Jasmine. 2009. *Armenian: Modern Eastern Armenian* (London Oriental and African Language Library 14). Amsterdam: John Benjamins. <https://doi.org/10.1075/loall.14>
- 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. Wiesbaden: Reichert.

- Grjunberg, Aleksandr. 1963. *Jazyk severoazerbajdžanskix tatov* [The language of the Tats of Northern Azerbaijan]. Leningrad: Izdatel'stvo Akademii nauk SSSR.
- Ġalt'axč'yan, Sergey. 1957. Hamařotaki Madrasayi hay-t'at'ereni masin [Briefly on Armeno-Tat of Mædræsə]. *Hayastan SSR Gitut'yunneri akademiayi teġekargir – Hasarakakan gitut'yunner* 4. 85–96.
- Ġalt'axč'yan, Sergey. 1970. Madrasac'inneri t'at'eren lezun [The Tat language of the residents of Mædræsə]. Yerevan: Unpublished Candidate of Sciences thesis defended at the Academy of Sciences of the Armenian SSR Sector of Oriental Studies.
- Ivanov, Vladimir & Lejli Dodyxudoeva. 2017. Sintaksičeskie otnošenija imen v severo-zapadnyx iranskix jazykax (na mazanderanskom i gilanskom materiale) [Syntactic relations of nouns in Northwestern Iranian language (based on a corpus from Mazandarani and Gilaki)]. *Voprosy jazykoznanija* 2. 77–95. <https://doi.org/10.31857/S0373658X0000971-7>
- Jügel, Thomas. 2016. Enclitic pronouns in Middle Persian and the placeholder construction. *Quarterly Journal of Language and Inscription* 1(1). 41–63.
- Jügel, Thomas. 2019. The development of the object marker in Middle Persian. In *Proceedings of the Eighth European Conference of Iranian Studies*, 203–219. St. Petersburg: State Hermitage Publishers.
- Karimi, Simin. 1990. Obliqueness, specificity, and discourse functions. *Linguistic Analysis* 20. 139–191.
- Karimi, Simin. 1996. Case and specificity: Persian *râ* revisited. *Linguistic Analysis* 3(4). 173–194.
- Lazard, Gilbert. 1970. Étude quantitative de l'évolution d'un morphème: la postposition *râ* en persan [Quantitative study of the evolution of a morpheme: the postposition *râ* in Persian]. In David Cohen (ed.), *Mélanges Marcel*, 381–388. The Hague: Mouton.
- Lazard, Gilbert. 1982. Le morphème *râ* en persan et les relations actancielles [The morpheme *râ* in Persian and grammatical relations]. *Bulletin de la Société de Linguistique de Paris* 77(1). 177–207.
- Lazard, Gilbert. 1994. Le *râ* persan et le *ba* chinois [The Persian *râ* and the Chinese *ba*]. *Cahiers de linguistique – Asie orientale* 23(1). 169–176. <https://doi.org/10.1163/19606028-90000382>
- Lopatinskij, Lev (ed.). 1894. Armjano-tatskie teksty [Armeno-Tat texts]. In *Sbornik materialov dlja opisanija mestnostej i plemen Kavkaza* 20, Section 2, 25–32.
- Mammadova, Nayiba. 2018. Eléments de description et documentation du tat de l'Apshéron, langue iranienne d'Azerbaïdjan. Paris: Institut national des langues et civilisations orientales dissertation. <https://tel.archives-ouvertes.fr/tel-01801775>
- Miller, Boris. 1929. Taty, ix rasselenie i govory (materialy i voprosy) [Tats, their distribution and dialects (corpora and issues)]. *Azərbaycanın əqrənən Cəmiyyətində Xəbərləri* 8(7). 3–33.
- Miller, Boris. 1945. Tatskie teksty: Materialy po govoram tatov Sovetskogo Azerbajdžana [Tat Texts: Materials on the Dialects of the Tats of Soviet Azerbaijan]. In Vasilij Abaev & Ivan Meščaninov (eds.), *Iranskije jazyki 1* [Iranian Languages 1] (*Iranica* 3), 107–126. Moscow: Izdatel'stvo Akademii nauk SSSR.
- Paul, Ludwig. 2008. Some remarks on Persian suffix *-râ* as a general and historical issue. In Simin Karimi, Vida Samiiian & Donald Stilo (eds.), *Aspects of Iranian linguistics*, 329–337. Newcastle upon Tyne: Cambridge Scholars.
- Paul, Ludwig. 2017. The case system of Modern West Iranian languages in typological and historical perspective (with special reference to *\*râdi*). Paper presented at the *Seventh International Conference on Iranian Linguistics*. Moscow: Lomonosov Moscow State University.

- Poplack, Shana & Stephen Levey. 2010. Contact-induced grammatical change: A cautionary tale. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and space: An international handbook of linguistic variation* 1, 391–419. Berlin: de Gruyter.
- Rastorgueva, Vera & Džoj Ėdel'man. 1982. Giljanskij, mazanderanskij (s dialektami šamerzadi i velätu) [Gilaki, Mazandarani (including the Shamerzādi and Velätu dialects)]. In Vasilij Abaev, Mixail Bogoljubov & Vera Rastorgueva (eds.), *Novoiranskije jazyki: zapadnaja grupa, prikaspijskie jazyki* (Osnovy iranskogo jazykoznanija 3), 447–554. Moscow: Nauka.
- Shokri, Guiti. 2018. Ditransitivity constructions in three Caspian dialects. In Agnes Korn & Andrej Malchukov (eds.), *Ditransitive constructions in a cross-linguistic perspective*, 135–145. Wiesbaden: Reichert.
- Soltanov, Ağacamal & Məmmədخان Soltanov. 2013. Tati-türki, türki-tati lüğət – Tatca-azərbaycanca, azərbaycanca-tatca lüğət [Tat–Azeri, Azeri–Tat Dictionary]. Baku: Qanun.
- Stilo, Donald. 2004. Coordination in three Western Iranian languages: Vafsi, Persian and Gilaki. In Martin Haspelmath (ed.), *Coordinating constructions*, 269–330. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.58.16sti>
- Stilo, Donald. 2005. Iranian as buffer zone between the universal typologies of Turkic and Semitic. In Éva Á Csató, Bo Isaksson & Carina Jahani (eds.), *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic*, 35–63. London: Routledge.
- Šokri, Giti. 1995 [1374]. *Guyeš-e Sāri* [The dialect of Sāri]. Tehran: Pažuhešgāh-e 'olum-e ensāni va motāle'āt-e farhangi.
- Windfuhr, Gernot. 1979. *Persian grammar: History and state of its study* (Trends in Linguistics 12). The Hague: Mouton. <https://doi.org/10.1515/9783110800425>
- Windfuhr, Gernot & John Perry. 2009. Persian and Tajik. In Gernot Windfuhr (ed.), *The Iranian languages*, 416–544. London: Routledge.

# Author index

## A

- Abbott, Barbara 115, 126  
Abdollahnejad, Elias 10, 15,  
205–206, 209–210  
Abolghasemi, Mohsen 6  
Adib Toosi, Mohammad Amin  
240, 247  
Adibifar, Shirin 94–96, 101  
Aghaei, Behrad 207  
Aijmer, Karin 64, 66  
Aikhenvald, Alexandra Y. 35  
Alizadeh, Ghazaleh 144–145  
Amrishahi, Mahshid 145  
Anand, Pranav. 23, 25  
Anderson, Cormac 48  
Anderson, Stephen 186, 191,  
210  
Annis, William S. 246, 251  
Anonby, Erik 6, 30–32, 34, 40,  
47–48, 50–51  
Apenchenko, V. S. 30  
Araz 38  
Authier, Gilles 277, 279,  
280–281, 284  
Azarmakan, Heshmatollah  
237, 240, 247

## B

- Babby, Leonard 223–224  
Bahar, Mohammad-Taqi 10  
Baltazani, Mary 271  
Baker, Mark 58, 103, 155, 159  
Bank, Sebastian 33, 39  
Barjasteh, Darab 144–145  
Barwise, Jon 22, 185  
Bateni, Mohammad-Reza  
8, 66  
Beaver, David 127, 132  
Benveniste, Emile 169  
Berger, Frauke 70  
Bernardos, Rebecca 3  
Bhatt, Rajesh 155

- Blažek, Václav 38  
Bopp, Franz 224  
Borjian, Habib 33, 43–44, 46  
Bork, Ferdinand 224  
Bossong, Georg 227  
Bostock, Mike 40  
Bouckaert, Remco 45  
Bowers, John 178–179  
Boyce, Mary 152  
Brame, Mike 157, 191, 197, 215,  
221, 231–232  
Brasoveanu, Adrian 125  
Browne, Wayles 155  
Bruk, S. I. 30  
Brunner, Christopher Joseph 166  
Bryant, David 36, 39, 45  
Bulut, Christiane 33–34  
Büring, Daniel 22  
Bynon, Theodora 138

## C

- Cathcart, Chundra Aroor 38  
Cereti, Carlo G. 152  
Chierchia, Gennaro 127  
Chomsky, Noam 8, 88, 158–  
159, 173, 199, 212  
Cinque, Guglielmo 162, 181,  
183, 186  
Citko, Barbara 207  
Coghill, Eleanor 138  
Cooper, Robin 22, 185  
Coppock, Elizabeth 127, 132  
Corbett, Greville 2, 90–91, 230  
Čunakova, Olga M 152, 140

## D

- Dabir-Moghaddam,  
Mohammad 7, 97, 119–120,  
125, 132, 145, 149  
Dalrymple, Mary 119–120, 132  
Darzi, Ali 65  
D'Arcy, Alexandra 64

- Den Dikken, Marcel 180  
Deo, Ashwini 237, 239, 244  
Diesing, Molly 160  
Diewald, Gabriele 66  
Dixon, R. M. W. 35  
Dodyxudoeva, Lejli 281  
Domínguez, Martín 31  
Downing, Bruce 207  
Du Bois, John W. 142  
Dum-Tragut, Jasmine 281–282  
Dunn, Michael 45

## E

- Eberhard, David M. 3  
Ēdel'man, Džoj 281  
Ehterami, Manoochehr 244  
Embick, David 58  
Emonds, Joseph 212  
Enç, Mürvet 124, 130, 160  
Erman, Britt 66  
Estaji, Azam 64  
Etesami, Parvin 240

## F

- Farkas, Donka F. 113, 121, 123,  
125, 130  
Fennig, Charles D. 3, 33–34, 38  
Finck, Franz Nikolas 224  
Fiset, Jean-Pierre 31  
Fodor, Janet D. 121  
Fonseca-Greber, BonniBeth 91  
Forkel, Robert 33, 39  
Forke, Diana 70, 73–77  
Franco Ludovico 180, 197  
François, Alexandre 35–38, 50  
Freeze, Ray 169  
Fuß, Eric 96

## G

- Galt'axč'yan, Sergey 227,  
280–282, 286  
Gawne, Lauren 62

- van Gelderen, Elly 66, 85–93,  
95, 100–101, 103
- Ghaderibaneh, Mohammad 210
- Ghanbari, Shahyar 241–242,  
251–252
- Ghaniabadi, Saeed 59
- Gheitasi, Mojtaba 32
- Ghomeshi, Jila 57, 59, 61, 65–67
- Gignoux, Philippe 152
- Givón, Talmy 87, 149, 201–202
- Gippert, Jost 38
- Göksel, Asli 211
- Golchin Arefi, Mandeh 145
- Grandjean, Martin 40
- Gray, Russell D. 36, 39, 45
- Grjunberg, Aleksandr 277,  
288, 290
- Greenhill, Simon J. 45
- Guéron, Jacqueline 169
- Gündoğdu, Songül 190–191,  
205–207, 209–210
- H**
- Hadjidas, Harris 96
- Haig, Geoffrey 2, 7, 33, 48, 78,  
85, 87, 94–99, 101–104, 181,  
190, 210
- Halle, Morris 58
- Hammarström, Harald 33, 39
- Han, Chung-hye 22, 25
- Hankamer, Jorge 257
- Harley, Heidi 168–169
- Haspelmath, Martin 32–33,  
39, 69
- Hassanpour, Amir 48
- Hayes, Amos 6, 29–31, 34,  
237–239, 244
- Heggarty, Paul 35–36, 48
- Heim, Irene 127, 130, 132, 160,  
179
- Höhle, Barbara 70
- Holmberg, Anders 160
- Hourcade, Bertrand 30
- I**
- Ivanov, Vladimir 281
- Izady, Michael Mehrdad 30, 33
- J**
- Jackendoff, Ray 212, 216, 218,  
220–221
- Janda, Richard David 62
- Jasbi, Masoud 9, 10, 59–60,  
107, 119, 127, 128, 132, 160, 277
- Jardine, Kevin 51
- Jaxontov, Sergei E. 38
- Johnson, Edwin Lee 6
- Jügel, Thomas 6, 33, 38, 43,  
47–48, 50, 93–97, 99, 137,  
139–141, 143, 145, 277, 292,  
294–295, 297
- K**
- Kahnemuyipour, Arsalan 61,  
92, 157, 181–183, 185–186, 190,  
197, 200, 269
- Kaiser, Elsi 15, 16, 19–20, 22, 24
- Kalyan, Siva 36, 50
- Karimi, Simin 8–9, 60, 119,  
122–123, 125, 132, 145, 155,  
157–158, 160–162, 166, 169,  
191, 197, 215, 221, 231–232, 261,  
266, 277
- Karimi, Yadgar 178–180, 186,  
200
- Karvovskaya, Lena 70
- Kaufman, Terrence 47
- Kayne, Richard 169
- Kazeminejad, Ghazaleh 145
- van Kemenade, Ans 85
- Kent, Ronald G. 6
- Kerslake, Celia 211
- Khanlari, Parviz Natel 6, 240,  
247
- Kim, Kyeong-min 22
- Kiparsky, Paul 237, 239, 244
- Kiss, Katalin É. 271
- Klima, Edward 17, 212,  
263–264
- Kolokonte, Marina 10, 257–  
258, 268, 270–272
- König, Ekkehard 74, 77
- Korn, Agnes 7, 43, 93, 297
- Kornfilt, Jaklin 58, 155–156, 159
- Kostinas, Ulla-Britt 66
- Krahmer, Emiel 132
- Kratzer, Angelika 132, 179
- Krause, Cornelia 208
- Kuno, Susumu 19
- Kwak, Saera 263
- L**
- Lambton, Ann K. S. 8
- Lapointe, Steven 186
- Larson, Richard 8, 59, 80, 173,  
179, 199, 210, 212, 214–216,  
218–220, 223, 225–226, 228,  
230, 233
- Lavine, James 207
- Lazard, Gilbert 8, 60–61,  
65–66, 76, 100–101, 145,  
167–168, 277–278
- Lecoq, Pierre 93
- Lehmann, Christian 201
- Lemey, Philippe 45
- Leskien, August 35
- Leung, Betty 22
- Levey, Stephen 283
- Lichtenberk, Frantisek 66
- Lopatinskij, Lev 277, 281–282,  
289
- López, Luis 261–262
- M**
- MacKenzie, David N. 43–44,  
48, 93, 98
- Maguire, Warren 36
- Mahdavi Mazdeh, Mohsen  
10–11, 167, 228, 237, 239, 251
- Mahmoodi-Bakhtiari, Behrooz  
2
- Mahmud, Ahmad 144
- Mahootian, Shahrzad 119, 125,  
127–128, 132, 155
- Mallick, Swapan 3
- Mammadova, Nayiba 277,  
284, 288
- Manzini, Rita 180, 197,  
206–207
- Marantz, Alec 58, 155, 159
- Matras, Yaron 45, 47
- McCloskey, James 207
- McConnell-Ginet, Sally 127
- McMahon, April 36
- Merchant, Jason 207, 257–258,  
267
- Miller, Boris 280–282
- Miller, Philip 186

- Mirtalae, Maedeh 237,  
247–249
- Mohammadirad, Masoud  
8, 32, 93, 97–98
- Mohebbi Bahmani, Hassan  
32–33
- Moorjani, Priya 3
- Moreau, Marie-Louise 34
- Moyne, John Abel 16–18, 25
- Murasugi, Kumiko 31
- N**
- Najafi, Abolhassan 237, 239
- Narasimhan, Vagheesh M. 3
- Nasser, Hayedeh 117
- Nejati Jazeh, Morteza 237,  
240, 247
- Nevis, Joel 186
- Nichols, Johanna 2
- Nikanne, Urpo 160
- Nikolaeva, Irina 119–120, 132
- Noyer, Rolf 58
- O**
- Öpengin, Ergin 48, 93, 97–98,  
100
- Oranskij, Iosif M. 43
- Oroji, Mohammad Reza 66
- Orywal, Erwin 30, 34
- P**
- Pakzad, Fazlollah 149
- Palmer, Adrian 181
- Parsipour, Shahrnush 146
- Partee, Barbara 116, 132
- Patterson, Nick 3
- Paul, Ludwig 5, 7, 43–44, 48,  
102, 277
- Payne, John R. 2
- Perry, John R. 3, 64–65, 288
- Pesetsky, David 218
- Pikkert, Peter 197
- Plank, Frans 224–225
- Poplack, Shana 283
- Potts, Christopher 116
- Preminger, Omer 155–156, 159
- Pulsifer, Peter L. 31
- R**
- Rasekh, Mohammad 101
- Rāsekhī, Vahideh 10–11, 171,  
257, 269
- Rasekhmahand, Mohammad  
144–145
- Rastorgueva, Vera S. 6–7, 281
- Reinhart, Tanya 18
- Reuland, Eric 18
- Rezaei, Amir 66
- Ringe, Don A. 45
- Ritter, Elizabeth 17–19, 22, 25
- Rizzi, Luigi 258
- van Riemsdijk, Henk 191, 212
- Roberts, Craig 66, 127
- Roberts, Ian 66
- Rohland, Nadin 3
- Rooth, Mats 265
- Roussou, Anna 66
- Runner, Jeffrey T. 15–16, 19–20,  
22
- Russell, Bertrand 115, 126
- S**
- Sabthematabadi, Parisa  
30, 34, 47–48, 50
- Sadeghi, Ali Ashraf 6
- Sag, Ivan A. 121, 257
- Salehi, Ali 210
- Samiian, Vida. 8, 59, 80, 157,  
173, 191–193, 197, 199, 202, 210,  
212, 214–215, 218–220
- Samvelian, Pollet 6, 95–97,  
137–138, 140, 145, 182, 186–191,  
195, 200
- Savoia, Leonardo 180, 197
- Schiering, René 86, 104
- Schmidt, Johannes 36
- Schmidtke-Bode, Karsten 201
- Schmitt, Rüdiger 30, 43
- Schnell, Stefan 97
- Schütze, Carson 80
- Schwarz, Michal 38
- Schwarzschild, Roger 115
- Sedighi, Anousha 8, 145–147
- Segal, Gabriel 179
- Shabani-Jadidi, Pouneh 8
- Shaked, Shaul 140
- Shamisa, Siroos 239, 244
- Sheyholislami, Jaffer 32–33
- Shokri, Guiti 281
- Siewierska, Anna 86–87, 91,  
102, 104
- Simons, Gary F. 3, 33–34, 38
- Simons, Mandy 127
- Singhapreecha, Pornsiri 180
- Skjærvø, Prods Oktor 6
- Sohng, Hong-Ki 22
- Soltanov, Ağacamal 290
- Soltanov, Məmmədxan 290
- Song, Jae Jung 80
- Stalnaker, Robert C. 126
- Stilo, Donald 5–7, 33–34, 44,  
46–47, 64–65, 68–69, 74–76,  
93, 226, 277, 296
- Storoshenko, Dennis R. 10,  
15, 22, 25
- Strawson, Peter F. 126
- Sussman, Rachel S. 15–16,  
19–20, 22
- Svenonius, Peter 212
- Šokri, Giti 281
- T**
- Tabibzadeh, Omid 237, 240,  
247–249, 253
- Taheri-Ardali, Mortaza 30–32,  
47
- Taleghani, Azita H. 8, 266
- Tanenhaus, Michael K. 15–16,  
19–20, 22
- Taylor, John R. 31, 63
- Tedesco, Paul 43
- Thiele, Hanna 96
- Thiesen, Finn 239, 242
- Thomason, Sarah 47
- Tenny, Carol 19, 24
- Todd, Terry Lynn 222
- Tonhauser, Judith 127
- Toosarvandani, Maziar 117, 243
- Torrego, Esther 218
- Torres Cacoullous, Rena 95
- Traugott, Elizabeth Closs 64,  
66
- Travis, Catherine 95
- V**
- Vahidian-Kamkar, Taghi 145,  
240–241
- Van Bogaert, Julie 67

- Vaughan, Jill 62  
Vinokurova, Nadya 58, 155, 159  
Vollmer, Maria 96
- W
- van der Wal Anonby, Christina  
45  
Whitman, John 58
- Windfuhr, Gernot 5, 30, 33–34,  
38, 43, 48, 65, 145, 277, 288  
Winkler, Susanne 261–262
- Y
- Yamakido, Hiroko 199  
Yoshie, Satoko 226  
Yousefian, Pakzad 48
- Z
- Ziamajidi, Leyla 247  
Zwicky, Arnold 186

# Languages index

## A

- Afghan Sign Languages 1
- aPersian 3, 5, 174
- Arabic 6, 45
- Aramaic 138
- Armenian 33–34, 275, 280–283
- Azerbaijan 9, 275–277, 290

## B

- Balochi 5, 7, 44–45, 49, 55, 174, 222, 296
- Bandari 34
- Bashkardi 34
- Basque 103
- Bushehr 32

## C

- Caspian 1, 5, 7–8, 44–46, 49, 173, 222, 226, 228, 230
- Central Kurdish 6, 48, 85, 93, 97–98, 100, 103, 205
- Central Plateau 44, 98
- Chahar Mahal va Bakhtiari 32
- Classical Modern Persian 155–156, 166–168 *see also* Persian
- Classical Persian 11, 145, 148, 237–244, 246–247, 249–251, 253–254
- colloquial Persian 11, 100–102, 107, 110, 116–117, 169, 237–238, 242–244, 246, 253–254

## D

- Dravidian 33

## F

- formal Persian 243, 254
- French 90–91, 151, 280

## G

- Georgian 103, 224
- Gilaki 1, 68–69, 93, 174, 222, 226, 228, 240, 254, 281
- Gorani 5, 44–45, 48–49, 174, 296
- Greek 96, 235, 244, 246, 250, 254, 271

## H

- Hawrami 5, 85, 93, 98, 174
- Hindi 90–91, 155
- Hormozgan 32

## I

- Ilam 32
- Indic 5–6, 33
- Indo-Aryan 33, 296
- Indo-European 3, 5, 33, 39, 45, 48, 275 *see also* Proto-Indo-European
- iPersian 3, 5, 8, 173–177, 181, 186–187, 189–193, 195–197, 199–202, 205–209, 211, 214–216, 218, 221, 226, 228–230, 232
- Iranian 1–9, 11, 33, 35, 39, 46, 50–51, 68, 78, 86, 91–95, 97–100, 102–104, 119, 137–138, 173–174, 190, 222, 240, 245, 275, 277, 283, 288, 296 *see also* Iranian, Middle Iranian, New Iranian, Old Iranian

- Iranic 3, 32–34, 38–39, 43–49
- Italian 206

## J

- Juhuri 275

## K

- Kartvelian 33–34
- Kordestan 32
- Korean 9–10, 16, 21–22, 80
- Kumzari 3, 34, 45
- Kurdish 1–2, 5–6, 8, 33, 44–45, 48–49, 54–56, 103, 151, 174, 178, 189, 209, 211, 221 *see also* Central Kurdish, Mukri Kurdish, Northern Kurdish, Kurmanji, Sorani
- Kurmanji 1–2, 174, 189–191, 197, 205–211

## L

- Lak of Kakevendi 98
- Larestani 34
- Latin 91, 244, 288
- Lori 5, 11, 34

## M

- Mandarin 16, 21, 23, 25, 233
- Mazandarani 1, 46, 93, 229, 281
- Mazanderani 174, 222, 226, 228–229
- Median 44
- Middle Iranian 3, 5, 94–95, 97, 99–100, 102, 138, 150
- Middle Iranic 33
- Middle Persian 3, 5–6, 48, 93–96, 139–142, 145, 149–151, 166, 169, 294–296
- Mukri 97, 100

## N

- New Iranian 138
- New Persian 7, 145–146, 151
- North Caucasian 6
- Northern Kurdish 93, 96–98, 102, 197, 205, 209



## O

- Old Iranian 33  
 Old Persian 3, 94, 99, 277  
 Old Persian 2–3, 5–6, 69, 145,  
 149, 166, 169, 277  
 Ottoman Turkish 242

## P

- Persian 1–3, 5–11, 15–17, 20–22,  
 24, 34, 40, 43, 48, 50, 57,  
 59–64, 68–69, 71–72, 74–75,  
 77–81, 92–97, 100–102, 104,  
 107–108, 110, 113, 116–120, 124,  
 127, 132–134, 138–142, 144–152,  
 155–157, 160, 162–163, 166–  
 169, 173–174, 180, 237–247,  
 249–251, 253–254, 257–258,  
 261, 263, 266, 268–269,  
 272, 275, 277–278, 285, 288,  
 294–296 *see also* aPersian  
 Classical, Modern Persian,  
 colloquial Persian, formal  
 Persian, iPersian, Middle  
 Persian, New Persian, Old  
 Persian

Proto-Indo-European 45, 69

## S

- Sakha 156, 159, 163  
 Semitic 5, 33–34, 45, 300  
 Semnani 5, 44, 48  
 Sign Languages 1, 33  
 Sistani 34  
 Sivand 93  
 Sorani 1, 137, 151, 174, 189,  
 205–206, 209, 210, 292  
 Southwestern Iranian 34, 45  
 spoken Persian 59, 96, 100–  
 102, 237

## T

- Tabaroid 46  
 Tajik 1, 5, 288  
 Taleshi 46  
 Talyshi 1, 5, 46, 222, 226  
 Tat 5, 9, 275–297  
 Tati 1, 5, 46, 275, 296  
 Tatic 44, 46  
 Tatoid 46  
 Tehrani 34, 107, 110, 113–118

- tPersian 3, 5, 174  
 Turkic 5–6, 33–34, 156, 159,  
 275–276, 288

## U

- Upper Sorbian 230  
 Uzbek 288

## V

- Vafsi 5–7, 68–69, 93  
 Vanuatu 36, 39

## W

- Wakhi 1, 5  
 West Iranian 38–39, 43–49  
 West Iranian 86, 93–94, 97,  
 102–104

## Z

- Zazaki 1, 5, 7–8, 44–45, 48–49,  
 93, 102, 174, 222–223, 225–226,  
 296

# Subject index

## A

- accusative alignment 137–138, 140
- agreement 6–7, 73, 78–79, 85–104, 142–145, 155, 181, 218, 224, 230–231
- agreement cycle 85–88, 90–92, 99
- agreement markers 6–7, 60, 74, 78, 86–87, 96, 98–99, 103, 137–140, 142–144, 148–151 *see also* topic agreement
- alignment 2, 7, 137–141, 233 *see also* accusative alignment, ergative alignment
- additive marker 8–9, 57, 59, 68–71, 73–79
- anacalasis 251, 254
- anacreontic 251
- anaphor 10, 15–16, 20, 22–23, 25
- anaphor resolution 15
- antecedent 10, 15–16, 19–25, 112, 115, 160–161, 259–260, 267, 272
- assignment function 22, 125
- attributive 175, 200, 201, 226, 230–232

## B

- benefactive 167, 169–170, 289
- beneficiary 144, 166
- bi-morphemic 18
- binding 10, 15–16, 19, 22–25
- binding domain 18
- bisyndetic coordination 69–70, 77
- bound variable 22–23, 25
- bridging context 7, 137, 139, 143, 150

## C

- Case marking 139, 150, 199, 221, 223–224, 268
- Case-assigning 199
- Case-assignment 159, 165–166, 199
- Case-marker 80, 155, 196, 225
- Case-marking analysis 59, 199, 201, 206, 210–212, 221, 225–226, 231
- Case-marking requirement 201
- Case-probe 230–231
- Case-requiring 199
- see also* double Case phenomenon, inherent Case, DP external Case, DP internal Case, non-Case-assigning, structural Case
- catalexis 246
- c-command 19, 24–25, 156, 159–166, 168–169
- Classical Persian poetry 237–244, 247, 249–251, 253–254
- classification, language 1, 3, 5–6, 29–33, 35–37, 39–40, 42–43, 45–51
- clausal projection 201
- clitic 9, 25, 57, 60, 66, 79, 85–87, 92–104, 138–139, 144–146, 148, 155–158, 161, 163, 170, 199, 201, 208, 210, 275–276, 278–279, 286, 289–290, 293, 295–296
- cliticization 7, 17, 86, 104
- clitic doubling 94, 100–101, 104, 155 *see also* enclitic
- co-indexation 19, 22
- comparative method 35–36, 39, 43, 46–47

comparative suffix 58

- complement clauses 8, 173, 200–201
- complex predicate 73–74, 144, 170
- compound verb 144–145
- concordiality/agreement 233
- concordializer 230–231
- Condition A 16, 18
- Condition B 16, 18–19
- contrastive focus 16, 95, 98, 258, 261, 265, 270–272
- coordination 10, 68–69, 77, 83, 90, 147, 257, 270–271
- co-predication 179–180
- co-reference 19, 23
- coreferent 97, 147
- cross-clausal referential relations 19

## D

- dative 141–142, 149–150, 155, 169–170, 224, 279–282, 284, 291
- de re blocking effect 24
- definiteness 10, 59, 61–62, 79–80, 92, 103, 107, 119–120, 125–127, 132–133, 155
- dependent Case 9, 60, 155–156, 159–160, 162, 165–166, 168–170
- derivation 9, 17, 25, 57–58, 61–63, 87–89, 115–117, 132, 134–135, 156, 178, 186, 231
- derived nominal 215–217
- Detachment Principle 67
- differential object marking 133, 145, 155, 275, 278
- directed graph 6, 29, 40, 44
- discourse markers 59, 63–64, 66, 79
- discourse particle 59, 63–68

- disjunctive Case hierarchy 159  
 double Case phenomenon 224, 225  
 doubled or strengthened Ezafe 222–226  
 DP external Case 223, 224  
 DP internal Case 223, 224
- E**  
 EDGE Features 186–187  
 ellipsis 10, 23, 25, 171, 257–258, 268, 271–272  
 emphatic 17–18, 22, 77, 90–91, 99, 103  
 enclitic 6–7, 9, 18, 57, 60–61, 67–68, 72, 78–79, 137–146, 148–151, 294–295  
 enclitic pronoun 7, 137–138, 140–143, 146, 149, 294–295  
 ergative 2, 103, 137–140, 142–143, 150–151, 159  
 ergative alignment 99, 137, 139–140  
 ergative subject 142, 150  
 experiencer 7, 137–138, 144–152, 276, 283–285, 297  
 experiencer construction 7, 137–138, 144, 148–151, 275  
 Ezafe 8, 17–18, 59–60, 62, 79–80, 148, 157, 173–178, 180–181, 183–197, 199–206, 208–212, 214–215, 218–223, 225–226, 228, 230–232, 285–286, 288, 292–293  
 Ezafe and Relative Clauses 176, 189  
 Ezafe Recursion 208–209
- F**  
 Finite Clausal Complement 200, 202, 205  
 Finite Relative Clauses (FRC) 176, 189–191, 200, 205–209, 228–229  
 focus 16, 65, 68–69, 75–76, 79–80, 90, 101, 120–121, 259, 261–262, 265, 269, 271–272, 275 *see also* contrastive focus  
 folk song 10–11, 237–240, 242, 244–247, 251, 253–254  
 foot 246  
 feet 244
- force-directed graph 6, 33, 40, 44, 49
- G**  
 gender 89, 102–103, 222  
 genealogical inheritance 6, 29, 43, 50  
 genitive 149, 159, 186, 191, 202–205, 208, 210, 223–224, 230  
 glottometric diagram 36–37, 39, 50  
 grammaticalization 7, 66–67, 76, 79, 85–86, 88, 93, 96, 102–104
- H**  
 hanging topic 7, 138–139, 143–144, 146–148, 150–151  
 hanging topic construction 97, 138–139, 143, 150  
 Head Movement 88, 185, 190  
 Head Preference Principle (HPP) 66–67, 87–89, 92, 103  
 head pronoun 89–90, 99  
 historical linguistics 38
- I**  
 indexation 17, 19, 22–23  
 information structure 261, 272  
 information structural accounts 120–121, 125, 132  
 inherent Case 224  
 internal nominality 208–209, 211  
 Ionic dimeter 250  
 iota operator 116, 132
- L**  
 lambda operator 22  
 language atlas 29, 49–50  
 language contact 9, 32, 38–39, 43, 46–47, 50, 275, 283, 296  
 language identity 48  
 language mapping 29–31  
 Late Merge Principle 67  
 lexical constraints 220  
 licensing requirement 8, 173, 196, 199  
 local binding 16  
 logophor 24
- logophoric 23–24  
 logophoricity 15, 23–24  
 long distance binding 16
- M**  
 monomorphemic forms 18, 21  
 monomorphemic xod 16, 22  
 moras 238–239  
 morphological affix 174, 186  
 multi-dimensional language relation web 38, 40–42, 49  
 multigraph 40, 49–50  
 multiple-constraints framework 10, 16, 19, 24
- N**  
 negation 10, 76, 85, 100, 257–258, 263–266, 268, 271–272  
 Negative Stripping 10, 257–268, 271–272  
 NeighborNet 36–37, 39  
 neutral intonation 21  
 nominal complementizer 206  
 nominal gerund 215–216, 218, 220  
 nominality 206, 208, 221  
 nominalization 211, 214, 218, 220–221  
 nominalized P 218, 221  
 nominalized pP 219, 221  
 nominals 10, 58, 107, 116–117, 120, 123, 128, 132, 158, 173, 191, 193, 197–202, 205, 215, 223, 226  
 non-Case-assigning 199  
 non-coreferential reading 17  
 non-nominal complementizer 206  
 null pronoun 22  
 Nunalit Atlas Framework 31
- O**  
 object 7, 9, 17–18, 22, 40, 60, 85–86, 94, 99–104, 119–124, 127–133, 138, 145, 151, 155–157, 159–162, 166, 179, 184, 192, 207–208, 212, 215–216, 218, 223, 232, 261, 269, 275–276, 278–283, 297  
 object agreement 7, 86, 91–92, 101–104

- object marker 9, 17–18, 60, 119–124, 128–132, 145, 155, 279, 282–283, 297
- object pronouns 7, 86, 90–92, 99–104
- oblique Case marking 139, 223–224
- oblique marking 277, 287
- omission 139–142
- omitted subject 139
- P**
- participle 2, 62, 191, 209, 228–229
- partitive Qs 185
- person-neutral 25
- plural marker 9, 61–62, 80, 150
- plural marking 80
- poetic license 240–241, 245, 251–253
- Polarity stripping 10, 258, 265–267, 271–272
- polysemy 63–64, 77, 79, 83
- polysemous 9, 61, 68, 74
- postpositions 296
- pragmaticalization 66–67
- predicate 16–17, 22–23, 25, 67, 73–74, 78, 94, 100, 103, 132, 144–145, 156, 164–166, 168, 170, 178–201, 220, 231, 260, 276, 283–284, 288–289, 297
- predicate-forming operator 178, 180
- Predication Marker 178
- predication structures 180
- See also* Stative Predicates
- preposition 78, 156–157, 160–161, 168–170, 192–194, 196, 197, 215, 220, 223, 225, 231, 284, 293–296
- prepositional complementation 212
- prepositional constructions 293
- presupposition 119–121, 125–128, 130–132
- Presupposed information 9, 117, 132
- processing 22, 63, 87
- pronominal enclitic 9, 72, 74, 78–79
- pronominalization 18
- pronoun 6–7, 10, 15–24, 66, 73, 77, 86–104, 137–151, 163, 206–208, 280, 294–296
- pronominalization *see also* pronominal enclitics
- Pseudo-stripping 10, 258, 260, 268, 270–272
- Q**
- quantifier 22, 110–112, 116, 121, 123–125, 132, 174, 185
- quantitative meter 10, 247–248, 251
- R**
- Ramal Mashkool 250
- reanalysis 7, 101, 137–139, 150
- reduced relative clauses (RRC) 176–177, 189–191, 195, 200, 205, 209, 228–229
- reflexive 16–17, 19–20, 146–147
- resumptive pronouns 95, 207–208
- Reverse Ezafe 8, 174, 222, 226–232
- roll-up movement 181, 183, 234
- Romance complementizers 207
- rukun 244
- S**
- scope 67, 69, 108, 110–113, 115, 118, 123–125, 127, 188, 218
- semantic binding 23, 25
- semantic information 19
- song lyrics 11
- source/perceiver contrast 19–22, 24
- specificity 9, 113, 119–125, 127, 130, 132–133, 155, 160, 278
- Stative Predicates 220
- stress 59–62, 95, 100, 229, 237, 240, 247–249, 296 *see also* unstressed
- Stripping 10, 257–258, 260, 263, 265–268, 270–272
- structural Case 159, 165–166, 223–224
- subject 16–25, 73, 78, 80, 86, 88, 90–92, 94–99, 102–104, 132, 137–151, 156–159, 161–166, 168–171, 178–180, 207, 239, 258, 261, 271, 277, 284, 288
- subject agreement 73, 78, 86, 88, 90, 92, 94, 96, 99, 102–103
- subjecthood 145
- Suffixaufnahme 224–225
- superheavy syllable 239–240, 242, 244, 251
- syntactic binding 10, 19, 22, 25
- T**
- thematic role 19–20
- Theme 25, 76, 145, 147–149, 151
- topic 7, 75–77, 96, 119–120, 137–138, 141, 150, 160, 164, 170–171, 261, 265
- topic agreement 7, 137–138, 150
- see also* hanging topic and hanging topic construction
- U**
- Uninterpretable 88–89, 92–93, 158, 266
- Uninterpretable/unvalued feature 232
- unstressed 60, 90
- V**
- Valued feature 232
- verbal agreement 6–7, 87, 137, 143, 148
- Verbal complementation 212
- Verbal gerund 215, 217–218, 220
- vowel shortening 11, 238, 241–245, 251, 253–254
- VP ellipsis 23
- W**
- weak subject orientation 22
- Z**
- zero expression 96, 139
- Φ**
- (phi-)Φ-features 18, 89–90, 92–93, 99

This volume brings together selected papers from the first North American Conference in Iranian Linguistics, which was organized by the linguistics department at Stony Brook University. Papers were selected to illustrate the range of frameworks, diverse areas of research and how the boundaries of linguistic analysis of Iranian languages have expanded over the years. The contributions collected in this volume address advancing research and complex methodological explorations in a broad range of topics in Persian syntax, morphology, phonology, semantics, typology and classification, as well as historical linguistics. Some of the papers also investigate less-studied and endangered Iranian languages such as Tat, Gilaki and Mazandarani, Sorani and Kurmanji Kurdish, and Zazaki. The volume will be of value to scholars in theoretical frameworks as well as those with typological and diachronic perspectives, and in particular to those working in Iranian linguistics.

ISBN 978 90 272 0716 6



9 789027 207166

**JOHN BENJAMINS PUBLISHING COMPANY**