2018

# CONSTRAINTS ON IZĀFA IN SORANI KURDISH

Ali Salehi
*University of Kentucky*, alisalehi67@gmail.com
Digital Object Identifier: https://doi.org/10.13023/etd.2018.351

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Ali Salehi, Student

Dr. Fabiola Henri, Major Professor

Dr. Rusty Barret, Director of Graduate Studies

# CONSTRAINTS ON IZĀFA IN SORANI KURDISH

---

THESIS

---

A thesis submitted in partial fulfillment of
the requirements for the degree of Master
of Arts in the College of Arts and Sciences
at the University of Kentucky

by

Ali Salehi
Lexington, Kentucky

Co-Directors: Dr. Fabiola Henri, Professor of Linguistics

and Dr. Gregory Stump, Professor of Linguistics

Lexington, Kentucky
2018

ABSTRACT OF THESIS

# CONSTRAINTS ON IZĀFA IN SORANI KURDISH

This study examines the distribution and the status of the izāfa particle in Sorani Kurdish (Central Kurdish). It uses a corpus-based analysis to investigate the forms and the pattern of distribution of the izāfa particle in Sorani, a dominant dialect of Kurdish among the Western Iranian languages. The study details an investigation of the appearance of izāfa in various NPs using a variety of data mostly from the corpus but supplemented by the grammaticality judgments of native speakers. I show that next to parallel properties seen in other Western Iranian languages, Sorani Kurdish izāfa shows a form alternation. I examine the morphological status of the izāfa and other nominal morphological features in Kurdish as well as the sensitivity of izāfa form variation to specificity in Kurdish NPs. I argue that the differences and distributional incoherence of the izāfa within Sorani and across Western Iranian languages calls for a morphomic approach, which can be formally described using a constructional approach to grammar. The study focuses on the following questions:

1. What type of head does the izāfa mark?
2. What is the function of this marker?
3. What are the constraints on its distribution?
4. What are the syntactic and morphological rules governing its distribution?

Keywords: Izāfa, Sorani Kurdish, Western Iranian Languages, Morphosyntax, Construction-based Approach.

Ali Salehi
July, 16, 2018

# CONSTRAINTS ON IZĀFA IN SORANI KURDISH

By

Ali Salehi

Dr. Fabiola Henri
Co-Director of Thesis
Dr. Gregory Stump
Co-Director of Thesis
Dr. Rusty Barret
Director of Graduate Studies
July, 16, 2018

# Acknowledgements

I am immensely grateful for the tireless dedication of the many people who assisted me with this work.

To Fabiola Henri who went above and beyond to direct me to keep this work on the right track. Her time, patience, motivation, enthusiasm, and vast knowledge are indispensable. She encouraged me to improve this thesis beyond my expectations.

To Gregory Stump whose insights and exceptional recommendations helped me in developing and refining the argument and analysis with different points of view in the course of writing my thesis.

To Mark Lauersdorf for his elaborate comments and ideas to improve the analysis in my thesis and for taking his time to inspire me in meetings that helped me look into the analysis in different aspects.

I am thankful to all three professors, not just for assisting me with my thesis but for encouraging me since the very first day of my masters degree.

In addition, I thank Pollet Samvelian for her encouragement and inspiration to pursue different ideas in my thesis. Her broad knowledge of Western Iranian languages helped me to enhance many aspects of this work.

I also thank all the other professors at the department of linguistics at the University of Kentucky.

I am especially indebted to Ali Ghaderi, my principal informant for the Kurdish data, for his patient and enthusiastic contribution to this study.

Table of Contents

# List of Tables

# List of Figures

# Abbreviations

| | |
|---|---|
| ADJ | Adjective |
| ADV | Adverb |
| CIRCUM | Circumposition |
| COP | Copular |
| DEF | Definte |
| DEM | Demonstrative |
| EZF | Izāfa |
| F | Feminine |
| HAB | Habitual |
| IMP | Imperative |
| IND | Indicative |
| INDF | Indefinite |
| IPFV | Imperfective |
| M | Masculine |
| NEG | Negative |
| SG | Singular |
| PST | Past |
| PL | Plural |
| POST | Postposition |
| PREP | Preposition |
| PRES | Present |
| PROG | Progressive |
| PRO | Pronoun |
| PST | Past |
| REFL | Reflexive |
| SBJ | Subject |
| SUBJ | Subjunctive |

# Chapter 1

## Introduction

Among Western Iranian languages, Kurdish is a term covering the largest group of related dialects. Despite linguistic differences, this cover term reflects the ethnic, cultural and linguistic unity among the speakers. The Kurds mostly inhabit a contiguous area spreading into parts of southeastern Turkey, northwestern Iran, northern Iraq and northern Syria (Figure 1.1). It is estimated that Kurdish is spoken by twenty or thirty million Kurds (Haig and Matras 2002, Salih 2014, Sheykh Esmaili and Salavati 2013). Kurdish comprises three main sub-group dialects namely Northern Kurdish (Kurmānjī), Central Kurdish (Sorani and Mukrī), and Southern Kurdish. Northern Kurdish dialects distinguish gender and have an inflectional system for masculine nouns (Haig and Matras 2002), while the others do not. According to McCaurus (2012), Kurdish, as a whole, occupies an intermediate position between North-western and South-western Iranian dialects. The main Kurdish dialect considered in this work is Central Kurdish (Sorani).

Most Kurdish people are bilingual or multilingual, speaking other languages, like Arabic, Turkish or Persian (Farsi), in addition to Kurdish (Salih 2014). Because of their interaction with speakers of other languages over a long period of time, Arabic, Armenian, Persian, New Aramaic and Turkish have left traces in various dialects of Kurdish. (Haig and Matras 2002).

Under the control of different empires, Kurdistan was historically divided into North and South by the Byzantine and Islamic empires and into Northwestern and East by the Ottoman and Persian empires (Nerwiy 2012). After the first World War, Kurdistan[1] was divided among Turkey, Persia, Iraq and Syria. In diaspora, Kurdish people have lived in Armenia, Lebanon, Egypt, parts of Europe and North America recently.

[1] I am using Kurdistan as a reference to the contiguous geographical area where Kurdish is spoken that is currently found in various political states.

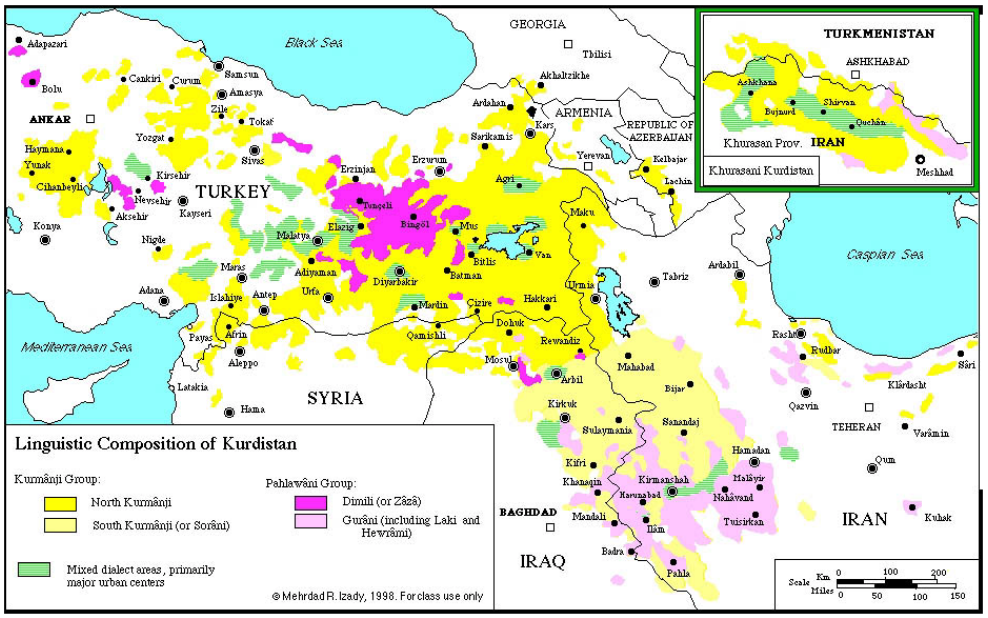


FIGURE 1.1: Linguistic Composition of Kurdistan (KAL 2018)

Kurdish is a language whose definition has caused some heated debates (Haig and Matras 2002). Sheykh Esmaili and Salavati (2013) consider Kurdish a 'dialect rich language, sometimes referred to as a dialect continuum' while categorizing a Kurdish dialect as a separate language or a language with distinct dialects is still controversial. If Kurdish were truly a dialect continuum, then neighboring Kurdish language varieties would be mutually intelligible to each other as one progresses along the continuum. The dialects of Kurdish are distinctly different and are not mutually intelligible (Bergsma et al. 2012). Haig and Matras (2002) suggest that the two main dialects of Kurdish, Northern and Central, can be even considered and treated as different languages among other Western Iranian languages. The traditional comprehensive classification of Kurdish dialects recognizes three main dialects of Kurdish. Other studies only focus on Northtern and Central dialects since they account for more than 75% of native Kurdish speakers. (Sheykh Esmaili and Salavati 2013, Zahedi and Mehrazmay 2011)

Most studies consider Kurdish as consisting of three main dialect subgroups: Northern Kurdish (Kurmānjī), Central Kurdish (Sorani and Mukrī), and Southern. Northern Kurdish dialects group is predominantly spoken by the Kurds in Turkey and all the Kurds of Syria (KAL 2018). Northern Kurdish (Kurmānji) encompasses a total population of about 20 million speakers (Sheykh Esmaili and Salavati 2013). Most Kurds speaking Central Kurdish (Sorani) reside in Iran and Iraq (KAL 2018, Nerwiy 2012). Central Kurdish has about seven million native speakers (Sheykh Esmaili and Salavati 2013). Southern Kurdish speakers live in Iran and Iraq. There are roughly 1.5 million Southern Kurdish Dialects speakers

in Kurdistan (KAL 2018). Haig and Matras (2002) state that the position of some concord markers is fixed immediately after a verb in Kurmānjī while in Sorani they are clitics and the lack of such clitics in Kurmanjī accounts for a major difference between the two dialects. Furthermore, Central and Southern Kurdish use the definite marker *-aka* and *-aga* respectively which is absent in Northern Kurdish dialects (Sheykh Esmaili and Salavati 2013).

### 1.1 Kurdish Orthography

Kurdish employs four scripts namely, Latin, Perso-Arabic, Cyrillic and Yekgirtú, their use depends on the geographical, political and cultural factors (Malmasi 2016). The Central dialects have used the Perso-Arabic script since the 19th century in the Kurdistan region of Iran and Iraq (Haig and Matras 2002). Northern Kurdish dialects use a Latin script introduced in the 1930s which has been used in Turkish and Syrian Kurdistan as well as in the European diaspora (Haig and Matras 2002). Most recent research regards Latin and Perso-Arabic scripts as the predominant scripts in Kurdish recognizing Kurdish as a bi-standard language as regards orthography (Malmasi 2016, Nerwiy 2012).

### 1.2 The Issue

This work is an attempt to suggest an analysis of the distribution and the syntactic pattern of izāfa marker in Kurdish. In Kurdish NPs, nominal modifiers following the head nouns are linked to it by two liaison izāfa morphemes (1)-(2)[2]; these indicate dependency relations between the head and its dependents McCaurus (2012).

(1) *gamāl ī zel*
dog EZF big
'big dog'

(2) *gamāl a zel-aka*
dog EZF big-DEF.SG
'the big dog'

(Fieldwork data)

Taking ideas from a whole range of previous literature, such as Ghomeshi (1997), Kahnemuyipour (2003), and Larson and Yamakido (2005), the present

[2]All the examples in the present work are taken from the news corpus (Malmasi 2016) (see §1.3.2) unless stated otherwise.

analysis draws primarily from the analysis of Persian izāfa proposed by Samvelian (2007b). The purpose of this study is to propose a fine-tuned analysis of the distribution and constraints of izāfa in Kurdish—an analysis that can account for all the forms of izāfa and form part of a wider description of izāfa in Western Iranian languages.

Many Western Iranian languages (e.g. Persian, various Kurdish dialects, Hawrami, Zazaki...) have SOV sentence structure. Kurdish NPs are similar to each other in several aspects:

- Most NPs exhibit head-initial[3] word order: relative clauses, possessor NPs, prepositional phrases and modifiers follow the head noun. There are, however, cases in which the head noun is preceded by a determiner or adjective if the structure is lexicalized.
- Within NP, a possessor NP always follows any adjectives.
- All elements appearing between the head noun and the possessor NP are linked to the head noun and to each other by izāfa. In Kurdish, izāfa is realized in two forms, both of them unstressed vowels. As Samvelian (2007b) observes izāfa is an enclitic in Western Iranian languages.

Izāfa (from Arabic *iḍāfa* meaning 'addition', 'annexation', or 'suppletion') is a significant structural characteristic of noun phrases in Western Iranian (Kahnemuyipour (2003), Larson and Yamakido (2005), Samvelian (2007b)). The Persian izāfa has been the focus of various studies and is generally described as a linker that appears between head noun and its dependents (Samvelian 2007b). Various analyses have been proposed to account for its properties: it has been analyzed as a phrasal affix (Samvelian 2007b), a phonological linking vowel indicative of phrasing (Ghomeshi 1997), as a case marker (Samiian (1994); (Larson and Yamakido 2005)), a marker associated with movement (Kahnemuyipour 2003, 2014) and as a conjunctive head (Rebuschi 2002).

Izāfa varies in its manifestation varies across Western Iranian languages, and, even more widely in the other languages (including Urdu, Armenian, and some Turkic languages) that have historically had contact with Iranian languages. In Sorani Kurdish, the focus of this study, post nominal modifiers are said to be

[3]Kahnemuyipour (2014) mentions some examples of the head final superlative NPs that alternate with head initial constructions that include izāfa. Also, Samvelian (2007b) mentions head final constructions without illustrations. In Kurdish, the superlative follows the same pattern, which will be discussed in section (3.2).

linked by an izāfa morpheme of varying forms that indicates dependency relations between the head and its dependents (McCaurus 2012). The constraints on the distribution of the different forms noted by McCaurus (2012) remain undescribed. In this thesis, I will show that besides exhibiting properties parallel to those seen in other Western Iranian languages, Sorani Kurdish izāfa additionally exhibits an allomorphy that is seemingly sensitive to definiteness. I will examine the distribution of izāfa with respect to:

1. Syntactic categories that call for this particular morpheme.
2. The function of the categories triggering the appearance of izāfa.
3. Sensitivity to markers including the plural, demonstrative and the indefinite marker in Kurdish.

I will argue that izāfa is iteratively affixed to modified headed constructions and that its appropriate form is determined by constraints on agreement with definiteness.

## 1.3 Methodology

The data used in the analysis include elicitations from native speakers and data from corpus. To test some of the constructions in Kurdish, grammaticality judgment tests have been used as well.

### 1.3.1 Data Collection

The data investigated in this work mostly comes from a corpus of Sorani Kurdish (Malmasi 2016) and collected texts from Twitter posts and Google in Sorani Kurdish mainly in the Iraqi Kurdistan area. I have recited data from 3 native speakers of Sorani the Kurdistan province of Iran. I have elicited data by giving the native speakers English and Persian sentences to translate. Also, grammaticality judgment is used to confirm my native speaker intuitions.

### 1.3.2 Corpus

Only a few Kurdish corpora have been compiled and made available for linguistic study. Nevertheless, the number of Sorani Kurdish texts available online

have been rapidly increasing in the last decade, which provides a good source of data for building corpora. The corpus used in the present work is a news corpus from Malmasi (2016) intended to document the subdialectal differences in Sorani Kurdish. The text of the corpus is taken from news providers based in Iran and Iraq. For Sorani Kurdish in Iraq, data has been collected from the Rudaw and Sharpress online news websites. For Sorani Kurdish in Iran, the data is from the Sahar TV and Kurdpress news websites. Articles from these news sources has facilitated the creation of a corpus of about 200,000 sentences and about 6,000,000 tokens total. This corpus[4] is freely available online. I have also used Kurdish Tweets and Google to look for the frequency of some structures. There are also a couple of other Sorani Kurdish corpora available that are not fully accessible. Lemmatization is not done in any of the corpora and part-of-speech tagging is done manually by a team of linguists only in one corpus Kurdish Language Corpus[5] in which lemmas are not considered in tagging. That is, *kitēb* 'book' and *kitēb-aka* 'the book' are considered two unique word types in that corpus, rather than two tokens of the same word type.

- Profile of the Corpus used in the present work from Malmasi (2016):
  - Total sentence count in the corpus: 196,575 sentences.
    - Sorani in Iraq: 74,573 sentences.
    - Sorani in Iran: 122,002 sentences.
  - Total type count in the corpus: 350,815 types.
    - Sorani in Iraq: 171,249 types.
    - Sorani in Iran: 179,567 types.
  - Total token count in the corpus: 6,052,932 tokens.
    - Sorani in Iraq: 2,198,834
    - Sorani in Iran: 3,854,098
  - Sources of the corpus texts: News agencies as described in the paragraph above.
  - Dates of the corpus texts: 2016-2017

[4]This corpus is last accessed on [07/19/2018] at http://web.science.mq.edu.au/ smalmasi/resources/sorani

[5]This corpus is last accessed on [07/19/2018] at http://kurdish-corpus.uok.ac.ir/

For the theoretical analysis, the insights of Samvelian (2007b) have been exploited at length, in conjunction with thorough introspection to arrive at classification of the distribution of izāfa in Kurdish with a formulation in a constructional approach. The study of the distribution of izāfa can help to learn about the function of this peculiar syntactic construction. Following the detailed description of the data, I will provide an implementation of the analysis within Sign-Based Construction Grammar, henceforth SBCG.

## 1.4 Outline of the thesis

In Chapter 1, I have introduced the particular idiosyncratic framework that I have employed in the work. I have illustrated it with some examples, and have presented the idiosyncrasies (incoherent distribution, form alternation) it brings up.

In Chapter 2, I provide a thematically structured review of previous research on izāfa in Kurdish and Persian. Alternative approaches to the description and analysis of izāfa construction are examined, with particular attention to the issue of concord within noun phrases and to the syntactic status of izāfa.

Chapter 3 provides a sketch of the distribution of izāfa in Sorani Kurdish, focusing on the categories that condition the appearance of izāfa in NPs. Bare, definite, and indefinite nouns, possessives, adjectival heads, complemetizer phrases and prepositional phrases are considered in a survey of noun phrase types involving adjuncts and complements.

Chapter 4 begins with the cliticization or affixation of izāfa. The sensitivity of izāfa in form alternation to specificity of the phrase with weak and strong definites is analyzed. In the last part, the incoherent distribution of izāfa is accounted for through a morphomic approach.

Chapter 5 describes izāfa hosted by constructs in SBCG and provides syntactic and morphological rules for the appearance and attachment of izāfa.

Chapter 6 presents a summary of the analysis and includes recommendations for further study.

**Chapter 2**

# Previous work

The Persian izāfa construction has been examined through the lenses of different approaches in Western Iranian languages and is still the focus of heated debates. A variety of analyses have been proposed to examine izāfa construction in the interface of syntax and morphology. In a grammatical account, izāfa is generally considered to be semantically vacuous. Hincha (1961), Samiian (1983, 1994), Ghomeshi (1997), Kahnemuyipour (2003, 2014), Larson and Yamakido (2005), Karimi and Brame (2012) and Samvelian (2007b) have proposed a range of different analyses of izāfa based on different sets of assumptions. Lazard (1966) uses the term *joncteur* 'linker' to refer to izāfa. Ghomeshi (1997) considers izāfa as a 'linker' inserted in the phonetic form to indicate phrasing. She states that izāfa is not a morpheme or functional head. Kahnemuyipour (2003, 2014) considers izāfa as a marker associated with the syntactic movement of the noun realizing a strong feature, in the minimalist sense. The first detailed study of Persian izāfa within X-bar theory was done by Samiian (1983). The function for izāfa has been labeled differently by various linguists indicating the different viewpoints on this construction:

- Izāfa as a phonological linker, phonological form (PF), to indicate phrasing. (Ghomeshi 1997)
- Izāfa as a case marker. (Larson and Yamakido 2005)
- Izāfa as a marker associated with roll up movements[1] of the noun in a minimalist analysis. (Kahnemuyipour 2003, 2014)
- Izāfa as a conjunctive head. (Rebuschi 2002)
- Izāfa as a linker indicating subject-predicate inversion. (den Dikken 2006)

[1] See §2.3 for details about roll-up movements.

- Izāfa as a phrasal affix. (Samvelian 2007b)

Thackston (2006) considers a dual behavior for izāfa interacting with attributive adjectives in Sorani Kurdish. The first type of izāfa is the 'open adjectival izāfa' which appears as an unstressed vowel linking an attributive adjective to a noun preceding it that is a bare noun, an indefinite singular or an indefinite plural, as in (3)-(5), respectively.

(3) *nāma* ***y*** *drež*
letter EZF long
'long letter'

(4) *nāma=yēk* ***ī*** *drež*
letter=INDF EZF long
'a long letter'

(5) *nām=ān* ***ī*** *drež*
letter=PL EZF long
'these dogs'

(Thackston 2006)

Izāfa vowel changes from *ī* to *a* when the linked noun and adjective are in a definite construction or enveloped by the demonstrative in Kurdish, as in (6) and (7). Thackston (2006) calls izāfa construction with *a* the 'close izāfa construction'.

(6) *hotel a bāš=aka*
hotel EZF good=DEF.SG
'the good hotel'

(7) *am hotel a bāš=a*
this hotel EZF good=DEF
'this good hotel'

(Thackston 2006)

If the noun ends with an *a* in izāfa construction with *a* (close izāfa construction), the linking *a* is omitted (Thackston 2006), as in (8) and (9).

(8) *čāyxāna gawra=ka*
teahouse big=DEF
'the big teahouse'

(9) *am čāyxāna gawra=ya*
this tea-house big=DEF
'this good hotel'

(Thackston 2006)

Thackston does not offer any explanation for his choice of the terms *open* and *close* in these izāfa construction. Note that the same phrase may contain both open and close izāfa constructions, as in (10).

(10) {*čāyxāna gawra=ka*} *y sar* {*šaqām a sarakī-aka*} *y šār*
teahouse big=DEF EZF on street EZF main=DEF EZF town

'the big teahouse on the major street in town'

(Thackston 2006)

In a historical descriptive account of izāfa in Northern Kurdish (Kurmanji) and Persian, Haig (2011) models the origins and distribution of izāfa in Old Persian and associates it with Northern Kurdish and modern Persian. Although there is nothing known about the possible lexical source of izāfa, it can be traced to the polyfunctional grammatical elements of Old Iranian (Haig 2011). Haider and Zwanziger (1984) and Karimi (2007) assume the forerunner of izāfa to be a relative pronoun which has gone under phonological and functional erosion. Haig (2011) believes that the relative pronoun hypothesis is only attractive if we look at Persian or Central Kurdish but Northern Kurdish and Zazaki seem less compatible with this hypothesis. Based on previous work on Avestan and Old Iranian, Haig (2011) claims that the presumed ancestors of izāfa are not simply relative pronouns. He states that the ancestor language to Northern Kurdish shows a demonstrative/relative pronoun very similar to that of Old Persian, whose izāfa shows a mix of demonstrative and relative properties. In Northern Kurdish all the original functions observable in Old Persian have been retained, while in Modern Persian the functional categories of izāfa (nominals) are lost and the particle is not completely disappeared.

For Northern Kurdish, Haig (2011) divides the functions of izāfa two broad types: an adnominal linking function and a demonstrative/anaphoric function. In the adnominal construction, izāfa links a post-nominal modifier to the head noun in definite and indefinite clauses and the gender and number determine the choice of izāfa particle, as in (11)-(12), but the sensitivity to definiteness is ignored.

(11) *mal-a* *mezin*
house(F)-EZF.F big
'big house'

(12) *zilam-ê* *li ber* *derî*
man-EZ.M in fornt of door:OBL
'the man in fornt of the door'

(Haig 2011)

The demonstrative/anaphoric function of izāfa is used independently of the head noun. Unlike the linking izāfa, this type occurs outside the phrase in which its antecedent occurs as "the pronominal head of its own phrase" (Haig 2011). Haig (2011) considers this function of izāfa resembles the characteristics of the English pronoun *one*. In this function izāfa comes close to having the properites of a nominalizer in Northern Kurdish.

(13) *şev-ên* *zivistan-ê* *dirêj in* *yên* *havîne* *kurt- in*
night-EZ.PL winter-OBL long-COP.PL EZF.PL summer short-COP.PL
'The nights of winter are long, those of summer are short'

(Bedir Khan and Lescot 1986)

### 2.1 Izāfa as a PF

Ghomeshi (1997) provides an analysis of the Persian izāfa construction and and the internal structure of Persian NPs. Ghomeshi (1997) posits that izāfa is a phonological linker at the level of phonological form (PF) and does not have a morphological status. She considers izāfa to be inserted in phonological form to indicate phrasing because of non-projecting properties of Persian nouns. Given the inability of Persian nouns to give rise to maximal projection, Ghomeshi suggests that the function of izāfa vowel is to link non-projecting heads to the elements with which they form a constituent. Ghomeshi's analysis is based on two predominant assumptions. First, heads can be attached to each other in syntax without projecting since some $X^o$ categories are non-projecting in syntax. Second, whether a constituent is phrasal or not does not solely depend on the projection of the head, but on whether that head itself is selected by a projecting element. There are three significant hypotheses that Ghomeshi (1997) suggests for Persian NPs (adapted from Samvelian 2007b):

1. Persian nouns are inherently non-projecting. That is, they never appear with filled specifier and complement positions and an NP node cannot dominate any phrasal nodes under.

2. Even though Persian nouns are non-projecting, they may themselves be selected by a projecting head such as $D^o$.

3. izāfa vowel never attaches to a phrasal constituent with [+N, -V] features (i.e. nouns and adjectives). The elements occurring within izāfa domain are just bare heads whether they are nouns, adjectives or prepositions.

Providing a descriptive account on the attachment of izāfa on different elements in Persian, Ghomeshi suggests that Persian DPs have the internal structure schematized in Figure 2.1.

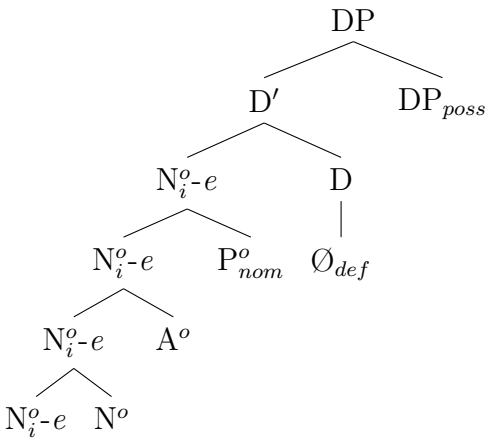


FIGURE 2.1: DP Structure

In Ghomeshi's analysis, the possessor DP is not dominated by any NPs but it is a part of the DP sister to the D′ which is a result of the hypothesis that Persian nouns are non-projecting. Based on the analysis, izāfa is inserted at PF on a lexical [+N] element that is followed by another independent constituent within the same extended projection (Ghomeshi 1997). In other words, Ghomeshi (1997) suggests that the function of izāfa vowel is to link non-projecting heads to the elements with which they form a constituent. Assuming that izāfa vowel does not correspond to any lexical or semantic feature and seems not to have any semantic features at all, Ghomeshi (1997) questions whether it makes sense to accord izāfa morphemic status. She suggests a rule for the insertion of izāfa vowel, operating in PF. The trigger for izāfa insertion is the feature [+N] borne by an X$^{o}$ followed by a modifier or complement.

Since izāfa appears before possessors in a phrase, Ghomeshi (1997) uses the notion of 'extended projection'. The idea is that N and D share the same categorical features with V and I but N and V are distinguished from D and I in their lexical status. Also, the insertion rule must account for the places that izāfa does not appear, that is, on the last X$^{o}$ within NP if the NP is followed by a definite determiner but no possessor. To solve this issue, Ghomeshi (1997) assumes that the rule of izāfa insertion is triggered only by phonetically overt material.

1. *Izāfa Insertion Rule*

   Insert the vowel *-e* on a lexical X$^{o}$ head that bears the feature [+N] when it is followed by phonetically realized, non-affixal material within the same extended projection.

Presuming that all of the lexical categories on which izāfa vowel appears are potentially non-projecting, Ghomeshi hypothesizes that the presence of izāfa identifies constituents that are not clearly given by the phrase structure hence the non-projection of [+N] elements in Persian is connected to, and part of the motivation for, the rule of izāfa insertion.

### 2.1.1 Problems with Ghomeshi's Account of Izāfa

There are a couple of empirical problems with Ghomeshi's analysis of izāfa. First, recall that the $N^o$ account of izāfa insertion in Figure 2.1 stipulates that izāfa marks lexical heads, either adjectives or nouns. This means that izāfa is not expected to mark already modified [+N] heads. But example 14 clearly shows that izāfa can appear on constructions featuring an already modified [+N] head.

(14) *pirāhan-e bozorg-e tāze*
shirt-EZF big-EZF new
'large new shirt'

(Fieldwork data)

In (14), the head is *shirt* and both adjectives are modifying *shirt*. Modification is triggering izāfa and the latter izāfa does not strictly need to appear on the lexical head; rather, it can appear on the modified head *pirahan-e bozorg*—a construction that is bigger than a simple lexical head. Second, Samvelian (2007b) posits that there are examples of phrasal adjectival modifiers which appear in an NP and are linked to the head by izāfa vowel, as in (15).

(15) *mard-e negarān-e bačče-ha-yaš vared šod*
man-EZF worried-EZF children-PL-PAF.3SG entered become-PAS
'The man worried about his children entered'

(Samvelian 2007b)

For Samvelian (2007b), this is evidence that there exist phrasal modifiers after izāfa occurring frequently in Persian.

## 2.2 Izāfa as a Case Marker

Samiian (1983, 1994) and Larson and Yamakido (2005) have the opposite view to Ghomeshi (1997). They consider izāfa as a case marker. Samiian (1983,

1994) provide detailed analyses of the Persian izāfa construction within an X-bar theoretical framework. Samiian mentions significant restrictions on the expansion of the constituents occurring within izāfa domain in Persian. According to Samiian, adjectives followed by a complement are excluded from izāfa domain and prepositions may take only nominal complements, while none of these restrictions hold outside izāfa domain. Samiian posits that the only recursively expandable constituent is the possessor NP. Samiian's study is significant because it is the first detailed study of Persian izāfa in a modern syntactic framework. Many of the works on izāfa in Persian follow the empirical facts mentioned by Samiian. This study suggests that constituents occurring within izāfa domain appear following a specific order and are constrained based on their distribution. In other words, different dependents to the head of the phrase connected by izāfa must appear in a fixed order. According to Samiian (1983), modifying elements linked to the head noun by izāfa must appear in the order in (16):

(16) Head N - Attributive Noun - Adjective Modifier - Prepositional Modifier - Possessor NP

(17) *ketāb -e tārix -e sabz -e bi arzeš -e Maryam*
book EZF history EZF green EZF without value EZF Maryam
'Maryam green history book without value'

(Samvelian 2007b)

Samvelian (2007b) suggests that the constraint on the position of the possessor NP is the real one. The possessor NP must appear in the final position within izāfa domain, any other positions are excluded for the possessor.

Samiian (1994) also considers izāfa as a case assigner appearing on nouns, adjectives and P2 and P3 prepositions(see (20)-(23)). Therefore, in Persian "anything that does not assign Case would need Case in order to be licensed and izāfa's function is to case-license [+N]" (Samvelian 2007b). The problem with this analysis is that case markers are usually assigned to arguments in a phrase not modifiers. Izāfa is promiscuous in its placement and is attracted to different categories and may appear several times within the same noun phrase.

Samiian (1994) also points out some major restriction on izāfa by a classification of the prepositions in Persian. Based on the analysis, three classes of prepositions exist in Persian:

- Prepositions that must not take izāfa (P1 prepositions): *dar* 'in', *tā* 'to, until', *be* 'to', *bi* 'without', *bā* 'with', *az* 'from'.

(18) *az* (**e*) koorosh
from (-EZF) Koorosh
'from Koroosh'

(19) *ba* (**e*) *kī*?
with (-EZF) who
'with who?'

(Fieldwork data)

- Preposition that must take izāfa, a large open class (P2 prepositions): *mesl-e* 'like', *dor-e* 'around', *baqal-e* 'by', *dāxel-e* 'inside', *pāin-e* 'below' .

(20) *dor* *(*e*) *mīz*
around EZF TABLE
'around table'

(21) *dāxel* *(*e*) *xāne*
in EZF house
'in the house'

(Fieldwork data)

- Prepositions that can optionally take izāfa (P3 prepositions): a small class that allow izāfa. Karimi and Brame (2012) consider this class a mixed class of prepositions: *tu-(ye)* 'inside', *ru-(ye)* 'on', *jelo-(ye)* 'in front of', *pahlu-(ye)* 'next to' , *bālā-(ye)* 'on top of' , *kenār-(e)* 'beside', *barā-(ye)* 'for'.

(22) *bālā* (*ye*) *dīvar*
on top of (EZF) wall
'on top of the wall'

(23) *tu* (*ye*) *otāq*
inside (EZF) room
'inside the room'

(Fieldwork data)

In line with Samiian (1994), Larson and Yamakido (2005) also consider izāfa in Persian as a case marker. To them, izāfa is a case marker that checks the case assigned by the head of a DP as its head to its arguments. Parallel to case assignment by ditransitive verbs in English, izāfa is assigned as a case marker to the constituents receiving it. In this analysis, most nominal modifiers are arguments of D. D and V both select thematic arguments so DP is parallel to VP. Within their analysis, nominal modifiers are considered to originate from D in postnominal position and DP modifiers are the lowest complements of the head beginning in post-head position. Since PPs and CPs do not have case features to be checked, they remain unmoved. Other modifiers that need case features to be checked are supposed to move to a site that they can check case such as the prenominal position. They consider constituents such as PPs and relative clauses excluded from izāfa domain since they do not need to be case marked. Since case marking

is associated with argument marking status, Larson and Yamakido use a 'shell structure' for DP based on Larson (1991). DP is similar to VP in taking advantage of a system of Case-marking.

- [+N] complements of D need Case - they bear a case feature that must be checked.
- D/$\delta$ can (in general) check Case on its internal argument, just as V/$\upsilon$ checks one Accusative on an internal argument of V.

These constraints result in the following:

- D will in general check Case on its NP restriction.
- DP-modifiers that do not have Case features to be checked (PPs, CPs, and disguised CPs) will remain *in situ*.
- DP- modifiers that bear Case features (APs) will be required to move to a site where they can check Case (e.g., by Concord)

(Larson and Yamakido 2005)

However, this analysis does not account for the combination of bare nouns with adjectives (bare N+Adj) and the combination of determiners with nouns (D+N) which occur frequently in Persian and Kurdish.

(24) *ketāw* **ī** *rash*
book EZF rash
'black book'

(25) *ketāw=aka*
book=DEF
'the book'

(26) *ketāw=im*
book=my
'my book'

(27) *ketāw* **ī** *min*
book EZF my
'my book'

(Fieldwork data, Sorani Kurdish)

In (25), the definite marker for singular *=aka* appears with a bare noun without any izāfa marking on the head. Also, in (24), the bare noun is appearing with izāfa marker without the presence of a determiner in the phrase. If the determiner is the Case assigner in (25), based on Larson and Yamakido's account, we should expect an izāfa to appear between the head noun and the specifier and *-aka* as the specifier of the phrase .

In addition, it is stated [+N] complements of D need Case, P is considered under N category based on in its distribution to account for the PPs that induce izāfa. Relative clauses as CPs do not require case and should appear in their base position without izāfa marker. Samvelian (2008) provides counterexamples to this prediction involving reduced relative clauses. The examples in (28) and (29) are from Kahnemuyipour (2014), with minor modifications.

(28) *in javān-e* [az suis bargašte]
this young-EZF from Switzerland returned
'this young man who has returned from Switzerland'

(29) *aks-e* [*čāp*-shode dar ruznāme]
picture-EZF published in newspaper
'the picture published in the newspaper'

### 2.3 Izāfa in a Roll up Movement

Kahnemuyipour (2003, 2014) examines izāfa in various word orders, registers and complex structures with an emphasis on its correlation with the word order of elements in the noun phrase. While the previous work (Ghomeshi 1997, Kahnemuyipour 2003, Samiian 1994) have claimed that the elements in izāfa domain are non-maximal, Kahnemuyipour (2014) states that the head status of the elements within izāfa domain is undermined once the dataset is expanded. Kahnemuyipour develops a phrasal analysis of izāfa construction using a roll-up movement.

In his work on the syntax and semantics of adjectives, Cinque (2010) develops an order in which the base order of the noun phrase is universally considered head final. Direct modification is lower in the structure of NP and indirect modification is higher in the structure. Any variation in the presumed order of NP is the result of phrasal movement in a roll-up fashion (See Cinque (2010), Kahnemuyipour (2014)). For example, in *beautiful beautiful dancer*, the first *beautiful* refers to the beauty of the person dancing (intersective) and the second *beautiful* refers to the beauty of dancing (nonintersective) (Kahnemuyipour 2014).

Kahnemuyipour adopts the roll-up analysis to the Persian izāfa construction. Izāfa is seen as a reflex of roll-up movement and the Persian DP is assumed to be head-final. NP is merged at the bottom of the tree structure and APs reside in the specifier projections above it. The demonstrative and the heads are higher up in the tree structure. The ordering of constituents that precede the head in a

noun phrase (such as English) is taken to be basic and post-nominal order (such as Persian) is derived from it. There are intermediate AgrP projections enabling the roll-up derivation, as in Figure (2.2). Kahnemuyipour suggests that under this view, izāfa marker can be seen as the surface realization of an inversion process, that is, a linker in the sense suggested by den Dikken (2006). He concludes that there is a correlation between the order of the noun and other nominal elements and the presence of izāfa marker in Persian, with the noun clearly demarcating the distribution of izāfa marker: There is no izāfa on elements surfacing before the noun and an izāfa for every element following it.

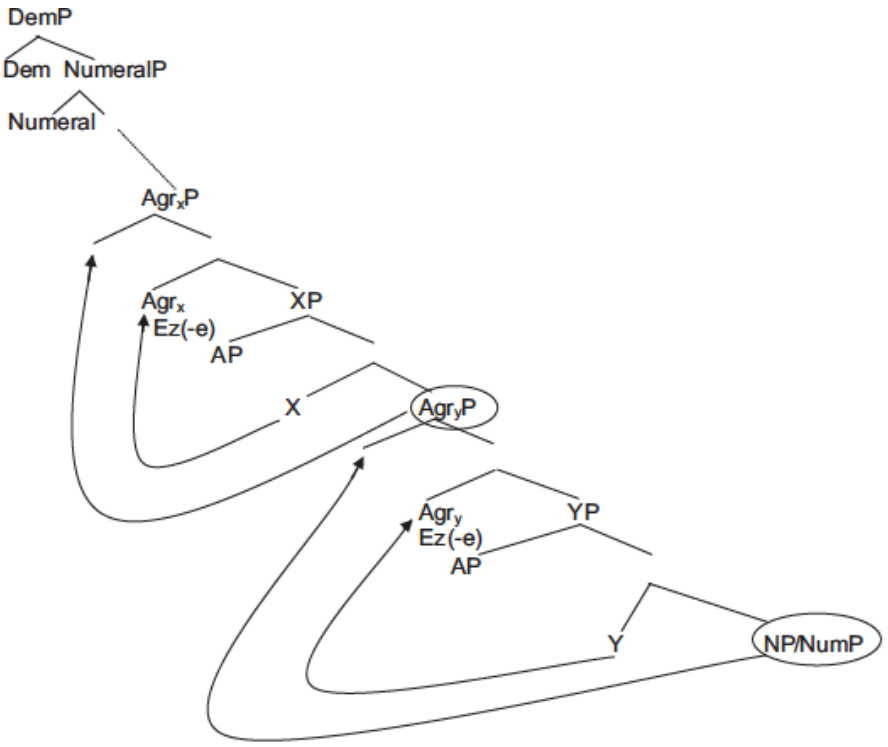


FIGURE 2.2: Deriving izāfa construction via roll-up movement (Kahnemuyipour 2014)

Kahnemuyipour posits that we cannot start with the assumption that these linkers are all the same element in all these languages, and pick properties from different languages to draw unified conclusions about them. Each language should be investigated with respect to all the properties of its NPs.

### 2.4 Izāfa as a Phrasal Affix

Samvelian (2007b) rejects the views of izāfa hypothesized in the previous works. Specifically, she rejects each of the following:

1. The assumption that izāfa is a Case-marker (Larson and Yamakido 2005, Samiian 1994)
2. The assumption that izāfa is a marker associated with the syntactic movement of the head noun recognizing izāfa marker can be seen as the surface realization of a linker (Kahnemuyipour 2003).
3. The assumption that izāfa is a conjunctive head (Rebuschi 2002).
4. The assumption that izāfa is a linker indicating subject-predicate inversion (den Dikken 2006).

Samvelian (2007b) points out the problems with Samiian's analysis of izāfa as a case-marker observing that the analysis is not based on well-grounded data. Also, Samvelian shows that constituents such as PPs and relative clauses are inside izāfa domain; this fact is problematic for both Samiian (1994) and Larson and Yamakido (2005), who exclude them from izāfa domain because they do not require case marking. Samvelian provides an affixal analysis of of izāfa on the basis of its distribution, morphology and prosody in the noun phrase. Two major sets of inflectional affixes are considered in Persian NPs. The first set of inflectional affixes attach to the nominal head of the phrase and are not separable by any other inflectional affixes. They bear the lexical stress and don't have a wide scope over the coordination of noun in the NP. The second set of inflectional affixes (including izāfa) are considered as phrasal affixes, that is- they appear at the right edge of nonmaximal nominal projections and are positioned after word-level affixes. Samvelian (2007b) proposes a dual behavior for izāfa, when it is attached to the nominal head, it functions as a word affix; but when it is added to a modifier, it functions as a phrasal affix.

Samvelian (2007b) provides a morphological account of izāfa in Head-driven Phrase Structure Grammar(HPSG), a constraint-based model of grammar. The scope of izāfa is always a word in this account. That is, a lexical suffixation rule is proposed with a function that takes a *plain-word* as input and produces a *phrasal-affix-word*. She observes that the immediate effect of a phrasal affix on the lexical

item it is attached to is to add a positive EDGE specification on the lexical item. Izāfa suffixation is done by two lexical rules in HPSG which are *word-to-word*. The first rule relates to the attachment of izāfa to the lexical head in an NP, AP or a PP when the scope of izāfa is always a word, as in (30).

(30) *in pesar-e-ye ahmaq*
this boy-DEF-EZF silly
'this silly boy'

(Samvelian 2007b)

Izāfa Suffixation Rule (I) (Samvelian 2007b)

(31) $$\begin{bmatrix} \textit{pl-nom-word} \\ \text{PHON } \boxed{1} \\ \text{CAT } \begin{bmatrix} \text{DEP} & - \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{pl-nom-word} \\ \text{PHON } \text{F}_{\text{EZ}}\,(\boxed{1}) \\ \text{CAT } \begin{bmatrix} \text{DEP} & + \end{bmatrix} \end{bmatrix}$$

Izāfa Suffixation Rule (I) in (31) is the input which takes in a *pl-nom-word* (plain nominal word) and returns an output where the *pl-nom-word* is suffixed with izāfa. Samvelian (2007b) suggests a Boolean right edge feature [+EZ] for the morphological rule of izāfa suffixation that can affix izāfa to words or phrases. When the rule applies to a [-EZ] word, it switches its specification to [+EZ]. The phonological form of the output is provided by the function $\text{F}_{\text{EZ}}$ as a morphological function that applies to a nominal base modeled as below.

TABLE 2.1: Ezafe Morphological Function (Samvelian 2007b)

| $X$ | $\text{F}_{\text{EZ}}(X)$ |
|---|---|
| $X$ ending with vowel $\neq$ [i] | $X$-*ye* |
| Otherwise | $X$-*e* |

The second izāfa suffixation rule applies only in noun phrases. This rule is used when izāfa links an N′ to a modifier or a possessor phras, as in (32).

(32) *mojgān-e az rimel sangin-e maryam*
eyelid.PL-EZF of mascara heavy-EZF Maryam
'Maryam's mascara-laden eyelids'

(Samvelian 2007b)

Izāfa Suffixation Rule (II) (Samvelian 2007b)

(33)
$$\begin{bmatrix} \textit{pl-wd} \\ \text{PHON } \boxed{1} \\ \text{EDGE} \begin{bmatrix} \text{EZ} & \textit{none} \\ \text{PAF} & \textit{none} \\ \text{INDEF} & \textit{none} \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{ph-af-wd} \\ \text{PHON } \text{F}_{\text{EZ}} (\boxed{1}) \\ \text{EDGE} \begin{bmatrix} \text{EZ} \begin{bmatrix} \textit{ph-af} & \\ \text{IN} & \text{N}' \begin{bmatrix} \text{DEP} & - \end{bmatrix} \\ \text{OUT} & \text{N}' \begin{bmatrix} \text{DEP} & + \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The second izāfa suffixation rule selects *plain-word* as its input and produces a *phrasal-affix-word*. The morphological function $\text{F}_{\text{EZ}}$ is used to adjoin izāfa. The values IN and OUT are attributes of the *ph-af* type which show the effects of izāfa as a suffix. The IN value takes in a constituent that must be N′[DEP −] and the OUT value provides the description of an N′ with dependents.

Based on the evidence for the morphological properties of izāfa and its interplay with other affixes of the same class such as personal affixes or the enclitic *-i* in Persian, the account of izāfa proposed by Samvelian (2007b) provides constraints on izāfa construction in Persian. Samvelian mentions that because of different grammaticalizations from Old Iranian to daughter Iranian languages, the process of change for *-hya* as the origin of izāfa particle in Old Iranian relative to the relative Iranian languages has presumably not proceeded to the same degree in all daughter languages giving birth to different realizations and forms of izāfa in various daughter languages.

Chapter 3 provides a sketch of the the distribution of izāfa in Sorani Kurdish focusing on the categories that induce izāfa in NPs. Bare, definite and indefinite nouns, possessives, adjectival heads, complementizer phrases and prepositional phrases are considered in a range of modified phrases with adjuncts and complements. The last part of the chapter deals with the types of prepositions in Kurdish and their interaction with izāfa. Most of the data is taken from the corpus of Sorani Kurdish (Malmasi 2016) with some data elicitation and judgment from native speakers of Sorani.

**Chapter 3**

# What is Izāfa?

The sentence structure in many Western Iranian languages, including Sorani Kurdish, is SOV, and hence head-final. However, NPs in Sorani are canonically head initial with a few constructions such as the superlative where the head noun is in final position. Definiteness, person and number are marked inflectionally, either through affixation or through cliticization. Izāfa (meaning 'addition', 'annexation', 'suppletion' in Kurdish) is of significant importance in the structure of Western Iranian languages. It is an unstressed grammatical particle which recursively appears between a head and its dependents. In Sorani Kurdish, the izāfa particle strictly appears in NPs and attaches to heads with a non-empty list of dependents. These dependents can be of various categories and assume different functions—specifiers, adjuncts or complements. I will show that the Sorani izāfa exhibits three allomorphs: the alternation of ***y*** [j] with ***ī*** [i:] is phonologically conditioned,[1] while their alternation with the third allomorph ***-a*** [æ] is seemingly sensitive to definiteness, more precisely specificity. Given the information we have about izāfa in NPs, the following questions need to be answered:

1. What are the syntactic categories that call for this particular particle? That is, Which lexical items can host this particle and which do not allow it?

2. What are the functions of the categories that trigger izāfa.

   Given that izāfa is always attached to the first element before the post-modifiers, What are the functions of the post-modifiers that trigger izāfa?

3. Are the analyses proposed in previously in line with corpus data and with the frequency of the izāfa's appearance?

4. Is izāfa represented by distinct allomorphs in NPs with demonstratives?

[1] Izāfa appears as -***y*** when the word ends with a vowel while it appears as ***ī*** when the word ends with a consonant.

5. How can izāfa be represented in the feature-based definition of NPs in a grammatical framework such HPSG or SBCG?

## 3.1 Nominal heads

In Sorani, nouns can be used bare or can be marked for different morphosyntactic features such as (in)definiteness or number. As the evidence below shows, these features can appear as syntactically independent words, clitics or affixes. Grammatical gender is not distinguished in Sorani Kurdish (unlike Northern dialects). The case distinctions in nouns and pronouns have been mostly lost and are replaced by a complex system of person marking (McCaurus 2012).

### 3.1.1 Bare nouns

Bare Kurdish nouns, i.e. those without any affixes or specifiers, may express a generic sense of the noun both in subject and object positions (34)-(35). Bare nouns can also be used as predicational complements (36).

(34) *qawa rash a*
coffee black is
'coffee is black'

(Fieldwork data)

(35) *aw=a prōfēsōr a*
he=DEF professor is
'He is a professor.'

(Fieldwork data)

(36) *māsi xoa=m*
fish eat=1SG
'I eat fish.'

(Fieldwork data)

As previously mentioned, an izāfa particle appears when the noun is modified[2] by dependents. In (37)-(38), the modification of the bare noun by the adjective *tāl* induces the appearance of the izāfa ***ī*** linking the head noun to the dependent AP.

(37) *haz ba qawa **y** tāl aka=m*
love from coffee EZF bitter do=1SG
'I love black coffee'

(Fieldwork data)

[2]I here use the term 'modify' in the large sense, i.e., one that covers specification, modification and complementation.

(38) *sawza* ***y*** *kāł a-xow=am*
vegetables EZF raw HAB-eat=1SG
'I eat raw vegetables.'

(Fieldwork data)

An NP may contain more than one izāfa: each dependent induces its own izāfa marking on the head it modifies. Hence in (39)-(40), two izāfas appear in the structure since the head noun is modified by two or more APs.

(39) *krās* ***ī*** *sur* ***ī*** *drīž*
shirt EZF red EZF long
'long red shirt'

(Fieldwork data)

(40) *gamāl* ***ī*** *zil* ***ī*** *rash* ***ī*** *hār*
dog EZF big EZF black EZF fierce
'big fierce black dog'

(Fieldwork data)

Bare nouns can also appear with PP dependents, both as complements (41) and adjuncts (42).

(41) *kitēb larbāra* ***y*** *zmānnāsī*
book PREP EZF linguistics
'book on linguistics'

(Fieldwork data)

(42) *kitēb lasar řafa bčuk-aka*
book PREP shelf small-DEF.SG
'book on the small shelf'

But unlike AP dependents not all PP dependents mark the head they attach to with an izāfa. Whereas the only or first PP dependent induces the izāfa in (43) and (44) respectively, the second PP in (45) does not trigger the izāfa marking on the head node. In addition, PP dependents seem to induce the izāfa only optionally. I return in more detail to the properties of prepositions in §(3.3) below. It is however notable that PPs differ from APs in that there is an adjacency requirement between a lexical noun and its PP dependent for the izāfa to appear compared to APs which

can induce the izāfa on an already modified noun. The behavior is further borne out when the PP combines with other dependents of different categories.

(43) *pyāw* **ī** *ba* *kłāw=ēk*
man EZF PREP hat=INDF
'man with a hat'

(Fieldwork data)

(44) *pyāw* **ī** *ba* *kłāw=ēk* *labān* *bāłkōn=ēk*
man EZF PREP hat=INDF PREP *balcony*=DEF
'man with a hat on a balcony'

(Fieldwork data)

(45) *pyāw* **ī** *barz* *ba* *kolāw=ēk*
man EZF tall PREP hat=INDF
'tall man with a hat'

(Fieldwork data)

Persian shows a similar behavior (Samvelian 2007b). I concur with Samvelian in assuming that some PPs might be outside of the izāfa domain. In the case of PPs, the appearance of izāfa seems to be linked with adjacency and perhaps extraposition, which would explain why no izāfa appears on the head noun—that is, the non-adjacent PP is out of the scope of izāfa and does not trigger this particle.

Like PPs, relative clauses are peculiar with respect to izāfa marking. McCaurus (2012: 621-23) describes the rule for Sorani relative clauses as involving (1) the conjunction *ka* 'that, who' if the clause is a specific definite, (2) an izāfa *-ī* if the clause is a restrictive definite and (3) no overt connector, or *-ēk*. What he identifies as a restrictive definite is in fact the expression *aw=a-y* 'the one', which is marked by an izāfa in the apparent absence of a noun (46). It is notable that the izāfa here appears outside of the scope of the definite marker compared to the definite NPs with modifiers, that is, izāfa appears in the **ī** form after *aka* which closes the boundary of the definite. (*cf.* § 3.1.3). This expression is usually followed by a clausal complement.

(46) *aw=a-***y** *ka* *hāt* *lagal* *t-ā*
that=DEF-EZF that came with 2SG-POST
'the one that came with you' (McCaurus 2012)

Some scholars state that that the complementizer *ka* appears in complementary distribtution with the izāfa in Persian (Haig 2011). The data shows that

there are examples where both the complementizer and the izāfa appear when the noun is modified by a relative clause (47)-(48).

(47) *tirs* **ī** *ka darnačū=m šēt=m ī kerd*
fear EZF that fail=1SG crazy=1SG RA do.3SG.PST
'fear that I will fail made me crazy'

(Fieldwork data)

(48) *gandał* **ī** *ka min bin=im*
corruption EZF that 1SG see.PST=1SG
'corruption that I saw'

In fact, the distribution of the izāfa with dependent clauses shows some inconsistencies. As seen above, bare nouns are marked with an izāfa when followed by a CP, whether the CP is a noun complement clause (47) or an adjunct relative clause (48).

Noun-noun constructions are also possible in Sorani, as in (49)-(50). Here a distinction must be drawn between noun-noun modifications and noun-noun compounds. Example (49) is a typical noun-noun modification in Sorani, with the izāfa appearing between the two nouns. This construction is head-initial; for example, *a linguistics book* is a kind of book rather than a kind of linguistics. Noun-noun modifications are analyzed along with genitives in §3.1.5. But as previously mentioned, head-final NPs are available in Sorani. An example is (50), a noun-noun compound[3] that is head-final(referring to a kind of fish rather than to a kind of snake). Since word-formation proceeds before inflection, it might be argued that the izāfa does not appear in such noun-noun compounds because the izāfa is a type of inflection. But recall also that head-final NPs never show the izāfa marking in any event.

(49) *kitēb* **ī** *zmānnāsī*
book EZF linguistics
'linguistics book'

(Fieldwork data)

(50) *mār-māsī*
snake-fish
'eel'

(McCaurus 2012)

Thus, head-final noun-noun compounds contrast with head-initial noun-noun modification in that they do not show an izāfa.

[3]For an account of compound nouns in Persian see Kahnemuyipour (2014).

### 3.1.2 Indefinite Nouns

In Sorani Kurdish, the marker for the indefinite singular *-ēk/-yēk* is postposed to the head noun. It exhibits a shape alternation, with *-ēk* appearing after consonants and *-yēk* after all vowels, as in (51) and (52).

(51) *pyāw=ēk*
man=INDF
'a man'
(Thackston 2006)

(52) *dargā=yēk*
door=INDF
'a door'
(Thackston 2006)

Indefinite plurals are formed by adding *-ān* to the bare noun. It is *-ān* if the nouns end in a consonant and *w* or *ū* and it is *-yān* if the noun ends in the vowels *y*, *ē*, *ō* and *ā*, as in (53) and (54) . Singular nouns ending in *-a*, as in (53), have a coalescence with the similar vowel in *-ān*.

(53) *nām=ān*
letter=PL
'letters'

(54) *dargā=yān*
door=PL
'doors'
(Thackston 2006)

Like bare nouns, indefinite heads, whether lexical or constructional, are also recursively marked with the izāfa when modified by dependents of various categories, as in (55). PPs with indefinite nouns require adjacency for the izāfa to appear; for example, compare (56) to (216).

(55) *gamāl=ēk* **ī** *zel* **ī** *rash*
dog=INDF EZF big EZF black
'a big black dog'
(Fieldwork data)

(56) *pyāw=ēk* **ī** *ba* *kolāw=ēk*
man=INDF EZF PREP hat=INDF
'a man with a hat'
(Fieldwork data)

(57) *kitēb=ēk* **ī** *kūrdī* *lasar* *Linux* *Ubuntu*
book=INDF EZF Kurdish PREP Linux Ubuntu
'a Kurdish book on Linux Ubuntu'

Inflected nouns specified for the indefinite modified by a CP clause may or may not host the izāfa, as (58)-(60) show. In these cases, izāfa marking does seem to

exclude the presence of the complementizer *ka*, suggesting that they might be in complementary distribution.

(58) *tirs=ēk* (***ī**) *ače ka awa be-mred*
fear=INDF (*EZF) goes that 3SG SUBJ-die
'there is a fear that he will die'
(Fieldwork data)

(59) *bapē=y kāt št=ēk* (***ī**) *ka ha-ya ba šēwa=y zanjīra tēkał bun=ēk?*
over=EZF time thing=INDF (*EZF) that exist-is PREP form=EZF series interaction entity=INDF
'Something that exists as a series of interactions over time?'

(60) *kas=ēk=ī šāraza=y āy tī dabē=t*
person=INDF=EZF specialty=EZF IT has=3SG
'someone who has IT skills'

#### 3.1.2.1 Quantification and indefinites

Some quantifiers obligatorily select indefinite nouns as in the corpus examples in (61)-(64). For instance, *čand* 'a few/some', *hamū* 'every', *čī* 'what' and *har* 'each' require a following indefinite noun. They crucially require that the specified noun be in the singular. These quantifiers are syntactically independent specifiers and hence are dependents of the head noun. Since the izāfa constitutes morphology whose conditioning is structurally determined by the presence of a dependent, we might expect that it would show up with some determined nouns.

(61) *hamū* (***ī**) *kas=ēk*
every (*EZF) person=INDF
'every person'

(62) *har* (***ī**) *šyn=ēk*
each (*EZF) place=INDF
'each place'

(63) *čand* (***ī**) *parēzgā=yēk*
some (*EZF) province=INDF
'some provinces'

(64) **čand* ***ī*** *parēzgā=yān*
*some EZF province=INDF.PL
'some provinces'

(65) *āmār=ak=ān čī dał=ēn?*
statistics=DEF=PL what say=3PL
'what do statistics say?'

However, these quantifiers do not induce an izāfa when combining with the head noun. Interestingly, the combination of nouns with such quantifiers provides a different NP structure where the head noun appears in the final position, in contrast

with structures where the izāfa appears which are head-initial NPs. On the other hand, Thackston (2006) notes that the expression *ī zōr* 'much, many, very' always shows a preceding ***ī*** izāfa.

(66) *mewān=ēk* ***ī*** *zōr*
guest=INDF EZF many
'many guests'

(67) *xarj=ēk* ***ī*** *zōr*
expenditure=INDF EZF many
'high expenditure'

But in fact, the licensing of the izāfa is subject to a number of non-categorical constraints. In the case at hand for instance, one does find constructions with *zōr* that do not exhibit any izāfa. These constructions show that word order constraints are determining in the licensing of the izāfa. Indeed, when *zōr* appears in a head-final NP, the izāfa is not licensed as is the case with the quantifiers in (66)-(67). Note also that in the former, *zōr* quantifies a noun but in (68), it modifies an adjective as an adverb would. Syntactic category and function could then also correlate with the presence/absence of the izāfa.

(68) *zōr xōšhāl bū=m ba-mānawa=m lēra*
very glad become=1SG SUBJ-stay-1SG here
'really glad that I stayed here'

I will argue in §3.1.5 that possessive specifiers are also exceptional in inducing an izāfa.

### 3.1.3 Definite Nouns

Sorani Kurdish is also marked for the definiteness. Like the indefinite, the definite marker features a shape alternation sensitive to the final segment of the host to which it attaches. Singular nouns are marked with *-aka* if they end in a consonant or any of the vowels the vowels *u*, *ē* and *ī* as in (70). Otherwise, if the noun ends in *a*, the definite marker is instead *-ka*, as in (69).

(69) *dargā=ka*
door=DEF.SG
'the door'

(Thackston 2006)

(70) *gamāl=aka*
dog=DEF.SG
'the dog'

(Fieldwork data)

Like indefinite nouns, definite nouns can be pluralized via the attachment of the plural clitic *ān* which may induce coalescence of the final vowel of the definite

(71)-(72). Similarly, the definite appears in a postnominal position, as in examples (69)-(72).

(71) *gamāl=akān*
dog=DEF.PL
'the dogs'

(Fieldwork data)

(72) *nāma=kān*
letter=DEF.PL
'the letters'

(Thackston 2006)

But while the indefinite which is almost always adjacent to the noun, the definite may be separated from the head noun by adjectival modifiers. In this respect, it has the character of an edge marker such as the English possessive *'s*, which appears at the right periphery of a nominal construction.

(73) *gamāl* ***a*** *zel=aka*
dog EZF big=DEF.SG
'the big dog'

(Fieldwork data)

(74) *gamāl* ***a*** *rash=aka*
dog EZF black=DEF.SG
'the black dog'

(Fieldwork data)

As expected from adjunction of modifying APs, an izāfa appears between the head noun and the adjectives (73)-(75). Interestingly, the definite property of the noun or NP triggers a different form of the izāfa (***a***) in contrast with previously examined data. Each adjunction induces its own izāfa marking on the head construction (75).

(75) *gamāl* ***a*** *zel* ***a*** *rash=akān*
dog EZF big EZF black=DEF.PL
'big black dogs'

(Fieldwork data)

Our previous observations about the definite *=(a)ka* showed that this marker behaves like an edge clitic, hence closing the NP boundary. However, examples with a head noun inflected for the definite *=aka* selecting PP dependents show that the latter appear outside of the scope of the *=aka*. For instance, compare (78) to (79), where the modifying adjective appears within the scope of *=aka* (which thus induces the *=a* form of the izāfa) while the modifying PP which appear outside of the domain of *=aka* and thus takes the *ī/y* form of the izāfa when it appears.

(76) *rūdāw=akān=ī la panjā=k=ān=ī sada=y rābrdū-dā*
event=DEF=PL=3SG in fifty=DEF.PL century=EZF past
*rūdadan*
happen
'Its events happened in the fifties of the past century'

(77) *pyāw=(*a) ba=klāw=aka*
man=EZF PREP=hat=DEF
'the man with the hat'

(Fieldwork data)

(78) *vāyrōs=akān=ī wak HPV*
virus=DEF.PL=EZF PREP HPV
'viruses like HPV'

(Glosbe 2018)

(79) *vāyrōs=a bkuž=akān=ī wak HPV*
virus=EZF deadly=DEF.PL=EZF PREP HPV
'deadly viruses like HPV'

(Glosbe 2018)

This means that there is a constraint on definite nouns to exclude PP dependents from their scope and as we will see, CP dependents, hence explaining the ungrammaticality of (77). This is irrespective of whether the dependent PP is an adjunct or a complement. In addition, the marking of the dependency with the izāfa is inconsistent (76)-(79) vs (80)-(81). The ***ī/y*** form of the izāfa appears outside the scope of the definite compared to adjectives which appear within the scope of *=aka* and trigger the ***a*** form of izāfa.

(80) *pyāw=aka (*a/*y)* ba *klāw=ēk*
man=DEF.SG (*EZF) with hat=INDF
'the man with a hat'

(Fieldwork data)

(81) *pyāw a čāk=aka=(*y)* ba *klāw=ēk*
man EZF good=DEF.SG=(*EZF) PREP hat=INDEF
'the good man with a hat'

(Fieldwork data)

It is also notable that the distribution of izāfa after *aka/akān* seems to be complementary with the 3SG marker since they have the same form; compare (82) to (76) repeated here as (83).

(82) *mindāl=akān=ī nāw (hāwrī yān nāhaz)=ī mīšēl mōrpurgō*
kid=PL.DEF=EZF in (friend or devil)=EZF Michael Morpurgo
'the boys in Michael Morpurgo's "Friend or Foe" '

(Glosbe 2018)

(83) *rūdāw=akān=ī la panjā=k=ān=ī sada=y rābrdū-dā rūdadan*
event=DEF=PL=3SG in fifty=DEF.PL century=EZF past happen
'the events happened in the fifties of the past century'

As noted in above, with CP dependents of a definite noun, the izāfa is optionally induced, with the CP clause appearing outside of the domain of the definite. Native speakers tend mostly to use the demonstrative *aw/am=X=a* instead of the simple definite *=aka* while in the written form both the demonstrative *aw/am=X=a* and the definite *=aka* are common. When the izāfa appears with this type of dependency, it is again subject to the scope constraint previously mentioned for PPs. As witnessed in (93), the definite marker appears on the head noun despite its modification by a CP clause rather than appearing at the edge of the NP (85). In addition, an izāfa marks the dependency in some cases, i.e, it appears on a noun already marked by the definite. Since it is not within the scope of the definite marker it appears in its *ī/y* form. Finally, we should note relative clauses are usually introduced by the relative marker *ka* as in (84) and (86). With definite nouns, this marker apparently alternates with the izāfa as in (83).

(84) *aw řāstī=ya (*ī) ka min zōr hala-bū=m labāra-**y** zōr šit-awa*
that true-DEF *EZF that 1SG very mistake-COP=1SG about-EZF many thing-POST
'that is true that I was wrong about many things'

(85) **tirs ī ače ka awa be-mred=aka*
fear EZF goes that 3SG SUBJ-die=DEF.SG
'there is fear that he will die' (Fieldwork data)

(86) *pyāw=aka/am pyāw=a ka lasar jēgā=ka=y dā-nīštūa*
man=DEF.SG/this man=DEF that on bed=DEF.SG=EZF HAB-sit.3SG
'the man who's sitting on the bed'

(Glosbe 2018)

Another construction where the constraint on the relative scope of definiteness regarding the form of izāfa is overridden is the possessive which I detail in §(3.1.5).

### 3.1.4 Demonstratives

The demonstrative *am/aw...(y)a* 'this/that' has previously been described as a kind of discontinuous envelope that surrounds the head noun (Thackston 2006)[4]. The types of nouns 'enveloped' by the demonstrative are bare forms. McCaurus (2012) analyzes *am/aw* as the demonstrative which obligatorily calls for a type of 'liaison morpheme' *-a-*.

(87) *am pyāw=a*
this man=DEF
'this man'

(Thackston 2006)

(88) *am pyāw=ān=a*
this dog=PL=DEF
'these men'

(Thackston 2006)

The plural marker in (88) is enclitic to the head noun and precedes the definite marker in the demonstrative; this contrasts with (71), where the plural marker follows the definite marker *=aka*. The apparent reason for this disparity is hiatus. In the case of the definite, *=aka* + *=ān* results in *-akān*. Following a similar rule, one would produce *=a* + *ān* = **-ān* which, would be homophonous with the plural marker, hence deleting any form marking of the definite. In that instance, Sorani seemingly chooses a different strategy. In the appropriate phonological context, the plural marker either cliticizes, as in (*=akān*) and (*ān=a*) to the definite markers *=aka* or *=a*.

It thus differs from other clitics in exhibiting properties of an ambiclitic, i.e, one which can appear on either side of its host (Booij et al. 2000).

The demonstrative can specify lexical nouns (87)-(89) or modified nouns resulting in complex expressions appearing between the discontinuous morphemes in (89). It is also notable that in (89), the infinitives followed by a modifier are inducing the izāfa. Infinitives in Sorani and other Western Iranian languages are also used as nominals. As such they can be considered mixed categories (nouns and verbs). When they are analyzed as nouns, they receive the izāfa marking when they select a dependent.

[4]Thackston (2006: p. 10) describes it precisely as a morphological envelope.

(89) *bo twānīn* ***ī*** ***am*** {*dyārī kirdin* ***ī*** *jegā* ***y***
in order to enable EZF this clarification do EZF position EZF
*Mĭr Gawra*} ***ya***
Mĭr Gawra DEF
'in order to enable this clarification of Mir Gawra's position'

(Thackston 2006)

The ***a*** form of the izāfa also appears with the demonstrative marker. As such, demonstrative marking can be seen as a type of discontinuous marking which obligatorily calls for an alternating form of the definite enclitic *-a*: *am...a* 'this' or *aw...a* 'that' (90)-(91). Together they mark the right and left periphery of the NP respectively.

(90) *am gamāl* ***a*** *rash a*
this dog EZF black DEF
'this black dog'

(Fieldwork data)

(91) *aw gamāl=ān* ***a*** *rash a*
that dog=PL EZF black DEF
'those black dogs'

(Fieldwork data)

As noted above, both noun complement clauses (or complement relative clauses) and adjunct relative clauses are selected by nouns determined by the demonstrative (92)-(94). These optionally feature an izāfa. As noted earlier, it seems that *ka* is in complementary distribution with izāfa when CPs appear after the definite *aka* or the demonstrative. More data is needed to support this idea but it is out of the scope of this study because due to lack of data distinguishing different types of relative clauses ( e.g. restrictive vs non-restrictive).

(92) *aw pyāw=a* (**y*) *ka l-awē ay-bīn-ī*
that man=DEF (*EZF) that PREP-there HAB-see-2SG
'the man that you see there' (McCaurus 2012)

(93) *aw řāstī=ya* (****ī***) *ka min zōr hala-bū=m labāra-****y*** *zōr*
that true-DEF *EZF that 1SG very mistake-COP=1SG about-EZF many
*šit-awa*
thing-POST
'that is true that I was wrong about many things'

(94) *aw pyāw=a y aw īwāra=ya qisa=y da-kird zōr*
that man=DEF EZF that evening=DEF talk=SBJ.3SG HAB-did very
*sarinjrākēš bū*
interesting was
'The man who talked at that evening lecture was very interesting.'

(Glosbe 2018)

PPs that appear with NPs marked with the demonstrative optionally receive an izāfa marking. (95)-(97).

there is variability with CPs and PPs to induce the izāfa.

(95) *am lēkčun=a lanāw mēšk=im-dā-ya*
this metaphor=DEF in brain=1SG-POST-IS
'this metaphor in my head'

(Glosbe 2018)

(96) *aw qutābī=a la-kalkudā*
that student=DEF PREP-Calcutta
'that student in Calcutta'

(Glosbe 2018)

(97) *aw sē řōž=a=y la pārīs*
that three day=DEF=EZF in Paris
'those three days in Paris'

(Glosbe 2018)

### 3.1.5 Possession

Kurdish uses two constructions to express pronominal genitives: The possessive specifier can either be enclitic to the head noun as in (98) or appear as a syntactically independent pronominal determiner in post-nominal position as in (99). Parallel to (in)definite inflected nominals, nouns bearing a possessive enclitic do not appear with an izāfa.

(98) *nāw=im*
name=1SG
'my name'

(McCaurus 2012)

(99) *nāw* ***ī*** *min*
name EZF 1SG
'my name'

(McCaurus 2012)

But like quantifiers, independent postnominal possessive pronouns induce an izāfa marking on the possessed noun. This parallels possessive constructions where the possessor is a full NP, as in (100). Minor postnominal determiners such as *kē*, as in (101), also appear with izāfa similar to the possessive determiner *min* in (99). Thackston (2006) assumes that izāfa appearing in possessives like (100) is semantically equivalent to the English preposition 'of'. However, this izāfa not only appears in (99) with the possessive determiner but as we have seen in the previous sections, it also appears with other modifiers; which means that it cannot always be seen as a type of preposition.

(100) *berā* ***y*** *Ali*
brother EZF Ali
'Ali's brother'
'brother of Ali'

(Fieldwork data)

(101) *kitēb* ***ī*** *kē*
book EZF whose
'Whose book?'

(McCaurus 2012)

English nominal PP complements of the type exemplified in (102)-(103) are expressed using the same possessive construction involving an izāfa in Sorani.

(102) *kitēb* ***ī*** *zmānnāsī*
book EZF linguistics
'linguistics book'
'book of linguistics'

(Fieldwork data)

(103) *ganj* ***ī*** *am šar a*
young EZF this city DEF
'young (people) of this city'

(McCaurus 2012)

Two observations are in order here. First, the constraint on allomorphy (***-a-***) previously described is overridden here since only the *ī* form is allowed. As expected with bare or indefinite heads, the izāfa appears as ***ī/y*** as in (104). But despite the presence of the demonstrative and/or the definite (106)-(107), the form of the izāfa remains unchanged.

(104) *kitēb=ēk=**ī*** *zmānnāsī*
book=INDF=EZF linguistics
'a book of linguistics'

(Fieldwork data)

(105) *andāzyār=**ī*** *šārazā=**y*** *kārga=ka*
engineer=EZF expert=EZF factory=DEF
'expert engineer of the factory'

(106) *čawrī **a*** *ziānmand=akān* ***ī*** *jasta y*
fat EZF harmful=DEF.PL EZF body 3SG
'his/her body's harmful fat'

This means that in genitives the izāfa is not subject to agreement constraints with the definite. Second, recall the observation about the definite *-(a)ka*, which I claimed to be an edge clitic, closing the NP boundary. With definite NPs modified by adjectival dependents (*cf.* (73)-(74) above), the izāfa appears within the scope of the definite. However, we can see that in genitive constructions such as that exemplified in (106), the izāfa form ***ī*** linearly follows the definite *=akan*. This is a subject-determiner (or head-specifier) type of genitive construction with the specifier position occupied by a full NP which can be closed by a definite marker. Note also that generally speaking, genitives are usually interpreted as definites. This means that if the *a* form of the izāfa is one which is sensitive to definiteness, it is expected to appear in genitives but never does.

(107) *dāyk o bāwk* ***ī*** ***aw*** *mendāl-an-**a*** ***y*** *pešwaxta la dāyk*
mother and father EZF that child-PL-DEF EZF early from mother
*būna*
become
'mother and father of those children born early'

### 3.2 Adjectival heads

Samvelian (2007b) argues that contrary to previous assumptions, the Persian izāfa not only marks head nouns but it also marks head adjectives and prepositions. Sorani adjectives also share with nouns the property of hosting an izāfa when appearing with dependents. Adjectives can indeed host the marker as shown by (108)-(109) which are cases of a head noun *krās* modified by the intersective modifying construction (submodification) [*sur* ***ī*** *rušen*], with the adjective *sur* exhibiting itself izāfa marking due to the dependency relation with the modifying

adjective *rušen*. Note that in English, 'a light red shirt' can mean 'a light shirt that is red' or 'a shirt that is light red', but *rušen* in Kurdish only indicates the meaning *light* in terms of color. So, the phrase is not ambiguous in Kurdish as it is in English.

(108) *krās-ēk* **ī** *sur* **ī** *rušen*
shirt-IND EZF red EZF light
'a light red shirt'

(Fieldwork data)

(109) *sur* **ī** *rušen*
red EZF light
'light red'

(Fieldwork data)

(110) *ba rang* **ī** *šīn-ēk* **ī** *kāl*
with color EZF blue-INDF EZF light
'in light blue color'

(Glosbe 2018)

The izāfa also appears on adjectives licensing phrasal complements. In (111), the izāfa **ī** appears on the adjective but not in (112) where the adjective selects a complement clause:

(111) *xam* **ī** *awa*
worry EZF her
'worried about her'

(Glosbe 2018)

(112) *zōr xōšhāl bū=m ba-mānawa=m lēra*
very glad become=1SG SUBJ-stay-1SG here
'really glad that I stayed here'

Unequal comparative forms are made by affixation of the suffix *-tir* to the adjective as in (113)-(115).

(113) *kas-ēk čāi sawz da-xwāt-awa, awā čawri* **a**
person=INDEF tea green IMPF-eat-POST DEM.PRO fat EZF
*ziānmand-akān* **ī** *jasta* **y** *kam-tir da-ben*
harmful=DEF.PL EZF body 3SG less-COMP. IMPF-are
'anyone drinking green tea has less harmful fat in their body'

(114) *barz-tir la Ali*
tall-COMP than Ali
'taller than Ali'

(Fieldwork data)

(115) *bčūk-tir la Ali*
small-COMP than Ali
'smaller than Ali'

(Fieldwork data)

This suffix only appears with scalar unequal comparisons but not in scalar equality comparisons (116) nor in non-scalar comparisons (117)-(118).

(116) *haw bālā* ***y*** *Ali*
same height EZF Ali
'same height as Ali'

(Fieldwork data)

(117) *tak awsā farq=ī da-ker-d*
from then different=EZF HAB-do-3SG
'It was different from then'

(Fieldwork data)

(118) *wak nān-aka* ***y*** *hafta* ***y*** *pē-š*
same food-DEF EZF week EZF before
'The same dish as last week'

(Fieldwork data)

Notice that the comparative form of the adjective may appear in a comparative construction that features a marker *la* which can be seen as synonymous to *than* in English comparatives as in (119).

(119) *min la tō barz-tir=m*
1SG from 2SG tall-COMP=COP
'I am taller than you.'

(Fieldwork data)

But more importantly, inequality comparatives do not induce the izāfa. By comparison, equality comparatives exhibit this marking in scalar (116) and not in non-scalar equality comparatives (117). Structurally, they resemble genitive constructions: (1) They use the default form ***ī*/*y*** of the izāfa; which again means that the presence of a definite marker does not have any influence on the form of the izāfa (118).

The superlative is produced by suffixation of *-trīn* on the non-head element of the construction e.g. *čāk* in (120). Since it is basically the extreme end of the comparison, it can thus be seen as a type of unequal construction. Given our previous observation, we expect not to find an izāfa with superlatives. This is borne out as seen from (120)-(121).

(120) *čāk-trin* *šit-ān*
good-SUPER thing-PL
'best things'
(Thackston 2006)

(121) *xrāp-trin* *šit-ān*
bad-SUPER thing-PL
'worst things'
(Fieldwork data)

### 3.3 Adpositional Heads

Sorani Kurdish dialects have a rich class of adpositional collocations with a complex syntactic behavior (Samvelian 2007a). The first categorization of adpositions in Kurdish is that of Mackenzie (1961), which takes account of two classes of adpositions in Kurdish: simple and absolute prepositions. Samvelian (2007a) uses almost the same type of classification, which will be exploited further here. Following most of her assumptions about primary adpositions, Simple and Absolute adpositions, I also distinguish two subtypes next to non-primary adpositions which here comprise compounded adpositions, converted adpositions and prosodically independent non-absolute adpositions (non-compounds).

#### 3.3.1 Primary adpositions

Simple adpositions include, among others, prepositions such as *ba* 'in, at, by, to' and *la* 'in' or 'from' and postpositions such as *-dā* or *-awa* 'in'. As expected from the terminology, prepositions precede the arguments they select (122)-(123) compared to postpositions where the selected argument precedes the adpositional head (124). Simple adpositions are in this study adpositions that neither etymologically nor synchronically identified as compounds (Table 3.1).

(122) *ba zmān-akān-**ī*** *arabi u turki u inglizi*
in language-DEF.PL-EZF Arabic and Turkish and English
'in Arabic and Turkish and English languages'

(123) *la diž **ī** aw xān-a šerpanjayāna **y** přōstāt*
PREP against EZF that tumor-DEF cancerous EZF prostate

'against that cancerous prostate tumor'

(124) *la-aw-lā-awa*
from-that-side-POST
'from that side'

(McCaurus 2012)

Table (3.1) combines the classification proposed in Samvelian (2007a) and McCaurus (2012) for Sorani Kurdish adpositions. The distinction between simple and absolute adpositions is initially made by Mackenzie (1961). It crucially distinguishes between two allomorphs of the same adposition, whose distribution is morphosyntactically constrained, more specifically by the "clitic versus non-clitic (non-affixal) realization of the complement" (Samvelian 2007a). In sum, the absolute form of the adposition appears only with clitic complements. According to Samvelian " the simple variant does not bear lexical stress and undergoes proclisis, while the absolute variant is accentuated" (Samvelian 2007a: p. 237). I return to absolute forms in §3.3.2 .

TABLE 3.1: Sample of adpositions in Sorani (adapted from Samvelian (2007a))

| Adpositions in Sorani | | |
|---|---|---|
| Simple | Absolute | Gloss |
| *ba* | *pē* | 'in', 'at', 'by', 'to' |
| *la* | *lē* | 'in', 'from' |
| *-a* | *-ē*∼*-r-ē*(postvocalic) | 'to' |
| *-da* | *tē* | 'in', 'and' |
| *bē* | - | 'without' |
| *bō* | *(bō)* | 'for' |
| *tā* | - | 'until' |

Some examples of 'simple' prepositions are *ba*, *la*, *bē*, *bō* and *tā* (125)-(134). Samvelian (2007a) analyzes *lagal* as a lexicalized preposition of the primary type. Here, however, I classify prepositions like *lagal* with non-primary ones because as I show, its behavior with respect to the izāfa is like that of compounded or converted adpositions.

(125) *la* *čāk-akān*
from good-DEF.PL
'from the good ones'

(McCaurus 2012)

(126) *ba hamāhangi lagal hokūmat* ***ī*** *surīa*
in coordination with government EZF Syria
'in coordination with the government of Syria'

(127) *la tō*
from 2SG
'from you'

(128) **la=t*
from=2SG
'from you'

(129) *pyāw-akān* ***ī*** *bē tāwān*
man-DEF.PL EZF without crime
'the men without crime'

(130) *xorāk-akān* ***ī*** *wak gošt* ***ī*** *sur*
food-PL.DEF EZF PREP meat EZF red
'the food like red meat'

(131) *wakū min*
PREP me
'like me'

(McCaurus 2012)

Postpositions such as *-a* 'to'[5], *-awa* 'from' or *-dā/ā* 'in' are exemplified in (132)-(133). They are enclitics to their selected phrasal complements and obligatorily appear with other prepositions, more precisely *ba* or *la*.

(132) *ba xwēy-awa*
with salt-POST
'with salt'

(133) *ba bāzār-ā čū-yn*
by bazaar-POST go.PST-1PL
'we went by way of the bazaar'

(McCaurus 2012)

In addition, McCaurus identifies *ba* as a (semantically) multivalent preposition and *la* as a bipolar one meaning 'in' and 'from' which is disambiguated by means of a postposition, requiring two terms or arguments (126) but analyzes the combination of the prepositions with postpositions illustrated in (128)-(135) as circumpositions[6]. But while the mentioned postpositions strictly require the presence

[5] McCaurus (2012) considers *-a* as a postverbal allative clitic.
[6] McCaurus (2012) also analyzes cases such as *ba X-awa* as an instance of circumposition.

of *ba* and *la* to be licensed, the reverse is not true. This is shown in (132)-(133) above, where both *ba* or *la* appear without the aforementioned postpositions. A simpler analysis would be to say that postpositions such as *-awa* strictly cliticize to a PP argument, i.e. prepositional head + Phrasal argument.

(134) *la karkuk-ā*
in Kirkuk-POST
'in Kirkuk'

(135) *la karkuk-awa bō baɣdā*
PREP Kirkuk-POST to Baghdad
'from Kirkuk to Baghdad'

(McCaurus 2012)

(McCaurus 2012)

McCaurus (2012) classifies the allative *-a* 'in' as a postverbal clitic. Its peculiarity lies in the fact that this adposition *-a* is an enclitic to the verb rather than to its argument (136). The allative clitic *-a* is reminiscent of the work of Klavans (1985) on clitics in which the same clitic is prosodically enclitic but grammatically proclitic.

(136) *a-č-īn-a karkūk*
HAB-go-1PL-POST Kirkuk
'we are going to Kirkuk'

(McCaurus 2012)

### 3.3.2 Absolute forms

In Sorani Kurdish, some simple adpositions exhibit an allomorphic variant called the absolute, with either a true form distinction or syncretic simple and absolute forms. According to Samvelian (2007a), it is the type of selectional restriction of these elements that allows one to distinguish the two classes. Indeed, the complement of a preposition can be realized as a syntactic item (NP, PP, or independent pronoun) or a bound personal clitic/affix. Pronominal and verbal clitics and independent personal pronouns are listed in Table 3.2 and 3.3 respectively.

As noted in the previous section, simple prepositions can select phrasal or syntactically independent arguments while absolutes only select clitic pronominals. Thus, the allomorphic realization is conditioned by the clitic vs. non-clitic(non-affixal) status of the complement. For example, in the absolute form *ba min* will be *p-ēm*

TABLE 3.2: Personal Enclitics and Verbal Endings in Sorani Kurdish (McCaurus 2012)

| Personal Enclitics | | | Verbal Endings | |
|---|---|---|---|---|
| | SG | PL | SG | PL |
| 1 | *im/-m* | *-mān* | *im/m* | *-īn* |
| 2 | *-it/-t* | *-tān* | *-ī* | *īn/n* |
| 3 | *-ī/-y* | *-yān* | *-ēt/ē/ ∅* | *-in/n* |

TABLE 3.3: Independent Pronouns in Sorani Kurdish (McCaurus 2012)

| Independent Pronouns | | |
|---|---|---|
| | SG | PL |
| 1 | *min* | *ēma* |
| 2 | *tō* | *ēwa* |
| 3 | *aw* | *aw-ān* |

and *la awān* will be *lē-yān*. In (137), the absolute preposition *lē* can host only the clitic form of pronouns (3PL here) and *pē* in (138) can host only the clitic form of 2SG. *pē* and *lē* can also occur with infinitives (139).

(137) *lē=yān/*awān*
to=3PL
'to them'

(138) *pē=t/*[pē tō] da-lē-m*
to=2SG HAB-say.PRS-1SG
'I am telling you.'

(Samvelian 2007a)

(139) *pē kan-īn*
PREP laugh-INF
'to laugh'

(140) *pē=m*
for=1SG
'for me'

(McCaurus 2012)

This clitic argument itself can be cliticized to either the absolute adpositional form i.e. as a clitic to the adpositional head (141) or as a verbal ending to other categories (142), with the requirement that both the adposition and its argument be adjacent to each other (Samvelian 2007c).

(141) *lē=yān rōjbāš a-kā*
to=3PL good morning HAB-do.PRES
'he wishes them 'good morning"

(142) *rōjbāš=yān* *lē* *a-kā*
good morning=3PL to HAB-do.PRES
'he wishes them 'good morning"

(Samvelian 2007c)

Samvelian (2007c) analyzes the non-local realization of adpositional clitic arguments as a case of extraction. However, as clitics, they are also expected to show promiscuous attachment as is often the case with these categories. This is true when the absolute forms are used in long distance dependency contexts, e.g. relative clauses, where the 'extracted' complement of the preposition occurs in the main clause and has no local realization in the subordinate clause, as in (143).

(143) *bird=iy=a* *aw* *sōn=a=y* *xoy* *lē* *bu*
take.PST=3SG=TO that place=DEF=EZF himslef to be.PST
'He took him to the place where he himself had been' (Samvelian 2007c)

In this case, "the clitic complement of the preposition can attach to the constituent in the phrase that immediately precedes the preposition" (Samvelian 2007c) (141)-(142).

The allative enclitic (*-a*) to the verb typically licenses its argument in the post verbal position; hence not attached to the verb itself, when its argument is a pronominal clitic, as in (144)-(145). McCaurus (2012) posits that *-a* occurs on the verbs of movement in alternation with with the preposition *bō* (146)-(147) while the absolute form of *-a* as *-(r)-ē* is found with verbs of giving (148)-(149).

(144) *bird-īy-a* *aw* *šuēn-a*
take.PST.3SG-POST that place-DEF
'he/she took him/her/it to that side'

(145) *bird-īy-a* *lā=yān*
take.PST-3SG-POST side=3PL
'he/she took it to their side'

(146) *kay* *a-č-ī* *bō* *baɤā*
when HAB-go.PRS-2SG to Baghdad
'when will you go to Baghdad?'

(147) *kay* *a-č-īt-a* *baɤā*
when HAB-go-2SG-POST Baghdad
'when will you go to Baghdad?'

(McCaurus 2012)

(148) *bō min bi-da-rē*
for 1SG IMP-give-POST
'give it to me'

(149) *bi-m-da-rē*
IMP-1SG-give-POST
'give it to me'

(McCaurus 2012)

The preposition *bō* is an example of a preposition with syncretic simple and absolute forms since it may select either a clitic or non-clitic complement (150)-(151). *lagal* 'with' is classified among simple prepositions in Samvelian's work but a compound in McCaurus's work.

(150) *bō min bi-da*
PREP 1SG IMP-give
'give me(some)!'

(151) *bō=m*
for=1SG
'for me'

(McCaurus 2012)

In addition, the prepositions *bē* and *tā* do not exhibit any absolute variant, meaning that these forms are defective. This means that they would never select a clitic complement. The simple adposition *-a* alternates with an absolute form *ē*, also an enclitic. Like its simple counterpart, it appears post-verbally (Samvelian 2007c). Absolute prepositions are stress-bearing compared to simple adpositions except for the enclitic preposition *-ē*.

### 3.3.3 Non-primary adpositions

Following Samvelian (2007a), I distinguish primary adpositions from compounded or converted forms. As noted throughout the literature, Sorani often compounds two adpositions. The first part of the compound is occupied by the prepositions *ba*, *la* or *a* while the second part is usually occupied by a converted nominal or adverbial that serves as an adposition (Samvelian 2007c). For example, in (152) *ba* is combined with *bē* to make the compound form of 'without'; in (153), *la* is combined with $nāw_{\text{N}}$ 'inside' to make the preposition inside; in (154) *la/ba* is combined with $dwāy_{\text{ADV}}$ to make the preposition 'after'.

(152) *anjāmdān ī aw kār-a* ***babē*** *twānā y zanestyān-**a***
do EZF that job-DEF without capability EZF science-POST
'doing that job without scientific capabilities'

(153) ***lanāw*** *aw bašdārbuw-ān-a y tir-da*
among that participant-PL-DEF EZF other-POST
'among those other participants'

(154) ***la-dwāy*** *rā kerdeni wušter-akān xāwan-akān=īan ba*
in order to make move do camel-DEF.PL owner-DEF.PL=3PL PREP
*otomobil-akān=īān-awa* ***ba-dwāy=an-da*** *rē-da-kan*
automobile-DEF.PL=3PL-POST PREP-after=3PL-POST move-HAB-do
'in order to make the camels run, the owners move after them with their automobiles'

There are also forms that feature the preposition *la* as the second part of the compound. While we saw that *la* can appear as a preposition alone, the corpus (Malmasi 2016) shows that the form *bējga* seems to always require the presence of the former to be licensed. Hence the analysis of the form *bējga la* as a compound.

(155) *hamū bējga la min*
everyone except from 1SG
'everyone except me'

(TKD 2018)

Samvelian (2007c) notes that it is not clear whether there are cases where it might be difficult to determine whether we have at hand true compounds or fully lexicalized forms. For example, Samvelian identifies the form *lagal* 'with' as a lexicalized form and groups it with primary prepositions as opposed to *la sar* 'on top of, according to' which would be a true compound since it combines the preposition *la* with a noun *sar* 'head'. She eventually analyzes all as lexicalized forms based on their selectional properties — selection of clitic argument vs non-clitic argument. I will here also consider these forms as lexicalized. However, I argue that despite its lexicalization, *lagał* is not synchronically analyzed as a compound, but as a type of preposition that behaves like compounds. This might have to do with the fact that they are non-monosyllabic or that they have diachronically evolved from compounded forms. *lagał* can also appear with a postposition as *lagał... (-dā)* 'together with'. in which *d-* is often elided or assimilated.

The third person personal enclitic can be homophonous with the izāfa (ī) but should not be confused with the latter. It is indeed in complementary distribution with izāfa and is often cliticized on the prepositions that usually select a 3SG (dative) argument (156)[7].

[7]See Samvelian 2007c: for more details.

(156) *twānyu-man-a lanēw šār* ***ī*** *karkuk dastgir-ī b-k-eyn u*
able-1PL-that in town EZF Kirkuk arrest-3SG IMPF-do-1PL and
*lēkolinawa lagal-****ī*** *bardawāma*
investigation PREP-3SG continuous
'we were able to arrest him in the city of Kirkuk and investigation about it is continued'

From a discriminative perspective, there are prepositions that look like compounds such as *jēgala* 'except for', *barla* 'before', *nēwān* 'between, among', *dagał* 'on, in, with', *labātī* 'instead of', among others. This is because like other adpositional compounds, they bear lexical stress; a property that they do not share with simple primary prepositions. Similar to both compounded and simple forms, they do select non-clitic arguments (158)-(159) but also clitic pronominals as in (157).

(157) *kuř=ēk dagał=mān la č=ıā-y birāndōst bu*
boy=INDF with=1PL from mountain-EZF birandost COP.PST
'a boy with us was from Birandost mountain'

(158) *paywandī dūstāna dagał drāwsē-kān*
contact friendly with neighbor-DEF.PL
'friendly contact with neighbors'

(159) *dagał tō*
with 2SG
'with you'

Note that whenever the second part of the compound has an absolute form, it is the latter that will be used with the clitic pronominal argument(160). In (161), the absolute form *pē* alternates with with *bē* when followed by a non-clitic argument.

(160) *nā-twān=em bapēy=ān bř̌ō=m*
NEG-can-1SG after=3pl go=1SG
'I can't go after them'

(161) *babē čāwdēr-****ī*** *pzīškī*
without care-EZF medical
'without medical care'

Non-primary adpositions also feature converts from other categories. For instance the forms *nāw* 'between, among' ≤ 'name$_N$', *pēš* 'before' ≤ 'forward$_{Adv}$',

*sar* 'above, on top of' ≤ 'head$_N$', *dwa* 'after' ≤ 'last$_{Adj}$', *balay* 'before' ≤ 'scourge$_N$' are originally converted from categories such as nouns, adverbs, adjectives, among others. The examples in (162)-(164) illustrate the use of some converted adpositions.

(162) *paul č-ū-a nāw āw=aka*
paul go-PST-POST in water-DEF.SG
'Paul got into the water' (Glosbe 2018)

(163) *balāy harīm-ī kordistān-awa*
next to region-EZF Kurdistan-POST
'next to Kurdistan region'

(164) *dwāy sāl=ēk*
after year=INDF
'after a year'

These converts share with other non-primary adpositions the property of bearing stress, which is unsurprising given their origin as a member of a lexical category. This again means that they have the status of independent words and their combination with their selected argument is purely syntactic, i.e. a head-complement type of relation. In investigating the corpus, we will see that converted adpositions behave like other non-primary adpositions with respect to izāfa marking.

### 3.4 Corpus Investigation: Keyword in Context

In order to get a full picture of the intricate behavior of adpositions in Sorani, I conducted a corpus analysis of raw data using KWIC (keyword in context). Since the distribution of the izāfa is the center of interest here, I examine the collocational window of words around it. Preprocessing for cleaning the corpus text (Malmasi 2016) to remove Latin characters and punctuation was done with R (quanteda package). The stopword list available in R for Kurdish could not be removed from the corpus since the list mostly consists of Kurdish prepositions. A window of 3-4 words was considered preceding and five words were considered following izāfa. Also, each preposition is considered with the same widow of words to examine the frequency of prepositions that may or may not host the izāfa. The reason for considering at least five words following each preposition is that there are instances of NPs in Kurdish with two or more PPs following in which, as mentioned earlier, only the more adjacent preposition to the noun can host

the izāfa. A list of 41 prepositions in Sorani Kurdish (McCaurus 2012, Samvelian 2007c, Thackston 2006) was used to investigate the attachment of the izāfa particle to each preposition. The frequency of each preposition is counted regarding its attachment to the izāfa particle as presented in Table 3.4. This gives insight to identify the distribution of izāfa and shows the frequency of this particle and its attachment to the prepositions. The corpus (Malmasi 2016) is divided into Iranian and Iraqi subcorpora based on the texts from news websites in the two regions. Therefore, one can identify whether the prepositions have different preferences of hosting izāfa in the Iraqi and Iranian areas. The search for prepositions is done independently in the two subcorpora that include texts from news websites in Iran and Iraq. As mentioned in §1.3.2, the total token count of texts from Iraq and Iran is different; hence, the frequency of the tokens is normalized per 2,000,000 tokens.

Based on the observations made regarding the distribution of the izāfa in the previous sections, it is expected that adpositions, which are argument selecting categories would exhibit the marking in a dependency relation. However, adpositions overwhelmingly show a preference for no marking. We do find instances of prepositions marked with an izāfa displaying the marking when selecting a non-clitic type of argument. Corpus data shows that simple adpositions do not induce any izāfa marking when combining with their arguments while non-primary adpositions that show a syntactic independence can optionally host the izāfa. Table 3.4 provides the collocational frequencies of the izāfa on prepositional heads in combination with their arguments.

The absence of izāfa marking with basic or primary adpositions is seemingly an effect of the type of dependency between the latter and their arguments. Since primary adpositions are clitics, their combination with their arguments is related to a morphological dependency rather than a syntactic dependency. As yet, the izāfa can seemingly be said to be an instance of contextual morphology encoding a syntactic dependency. Simple prepositions are proclitic, as Samvelian (2007a) noted, and hence are not expected to show an izāfa in combination with their selected argument. Postpositions would predictably not induce any izāfa either. There might be at least two different explanations for this behavior. If these postpositions are clitics, the prediction is that the izāfa would not appear. If we assume that postpositions combine with their arguments syntactically, the prediction is still the unmarked form since the construction would be head-final construction, which as previously noted never shows the izāfa.

TABLE 3.4: Frequency of Prepositions with/without izāfa in Sorani; Normalized per 2,000,000 tokens

| | Iraqi Sorani | | | Iranian Sorani | |
|---|---|---|---|---|---|
| Preposition | with izāfa | without izāfa | Preposition | with izāfa | without izāfa |
| *ba* 'for, with, by' | 0 | 332 | *pē* 'for, with, by' | 0 | 666 |
| | 0 | 122 | | 0 | 2,073 |
| *la* 'in, at' | 0 | unknown | *lē* 'in, at' | 0 | 696 |
| | 0 | unknown | | 0 | 511 |
| *babē* 'without' | 0 | 109 | *badam* 'along with, while' | 0 | 31 |
| | 0 | 493 | | 0 | 1 |
| *balāy* 'in the opinion of' | — | 13 | *barla* 'before' | 0 | 133 |
| | — | 101 | | 0 | 3 |
| *bapēy* 'according to, after' | — | 2,677 | *baraw(ī)(a)* 'in the direction of' | (*ī*)3-(*a*)28 | 1,164 |
| | — | 2,805 | | (*ī*)0-(*a*)8 | 653 |
| *bardam(ī)* 'before, in front of' | 10 | 297 | *bē* 'without' | 0 | 2,937 |
| | 20 | 245 | | 0 | 996 |
| *bējgala* 'except for' | 0 | 90 | *bō* 'for' | 0 | 17,637 |
| | 0 | 14 | | 0 | 28,771 |
| *dagaɬ* 'on, in' | 0 | 58 | *dwāy* 'after' | — | 2,248 |
| | 0 | 4 | | — | 4,556 |
| *jigala* 'except for, aside from' | 0 | 126 | *balay* 'next to' | — | 14 |
| | 0 | 436 | | — | 98 |
| *labar* 'because of' | 0 | 1,171 | *labābat* 'about, concerning' | 1 | 32 |
| | 0 | 663 | | 2 | 0 |
| *labāra(y)* 'about, concerning' | 746 | 8 | *labātī* 'instead of' | — | 6 |
| | 1,917 | 12 | | — | 0 |
| *labin* 'beside' | 0 | 9 | *lagal* 'with, together with' | 19 | 9,647 |
| | 0 | 16 | | 66 | 7,351 |
| *lalāyn* 'by (passive agent)' | 7 | 827 | *lanāw* 'within' | 3 | 220 |
| | 14 | 2,173 | | 1 | 268 |
| *lapēnāw* 'for the sake of' | 3 | 178 | *larēy* 'for, by means of' | — | 3 |
| | 33 | 138 | | — | 9 |
| *lanēw* 'between, among' | 0 | 333 | *larēgāy* 'for the sake of' | 2 | 0 |
| | 0 | 1,204 | | 8 | 0 |
| *lasar(ī)* 'on top of, according to' | 38 | 3,953 | *lažēr(ī)* 'under' | 0 | 171 |
| | 91 | 6,593 | | 1 | 484 |
| *nāw(ī)* 'between, among' | 631 | 384 | *nēwān(ī)* 'between, among' | 3 | 2,713 |
| | 1017 | 253 | | 8 | 1,966 |
| *pāš(ī)* 'after' | 3 | 1648 | *pēš(ī)* 'before' | 67 | 1,005 |
| | 2 | 513 | | 24 | 1,063 |
| *wak* 'like' | 0 | 1,230 | *labayn* 'midst, among' | 0 | 0 |
| | 0 | 1,783 | | 0 | 0 |
| *tā* 'until' | 0 | 2,073 | *dwā* 'last' | 0 | 251 |
| | 0 | 1,380 | | 0 | 258 |
| *tanīšt* 'next to, beside' | 1 | 115 | | | |
| | 2 | 37 | | | |

Absolute adpositions never host the izāfa either. Again multiple factors converge in explaining why that would be the case. Since they bear stress, one might think that they are independent words. Absolute prepositions are indeed accented which is an atypical property of clitics and their combination with their clitic argument allows the creation of a prosodically independent word. But even as syntactically independent words, these would still not be marked by an izāfa in combining with their arguments since their arguments themselves are clitics. The fact that they are the alternate forms obligatorily used with clitic arguments signals

their morphological dependency with their arguments. Based on the data examined above, it can be concluded that the izāfa is a type of contextual morphology that signals the structural relationship between a head and its dependent. While clitics are usually (re)ordered in syntax, my assumption is that izāfa is handled by morphology. Like other clitics previously described, the izāfa is a clitic whose phonological attachment is determined by morphology but whose distribution is constrained by syntax.

Non-primary adpositions on the other hand typically involve a syntactic relation and thus the dependency would predict the izāfa to appear between the adposition and its argument. Despite the fact that it does not systematically appear, this prediction is borne out. For example, *labāra* appears overwhelmingly with the izāfa, as in (165), with only a low number of instances of appearing without izāfa while *lagaɬ* generally appears without an izāfa. Table 3.4 shows that adpositions that can be identified as compounds, converts or adpositions with word-like status — in being stress-bearing, possibly being derived diachronically from compounds despite synchronic lexicalization or having a phonological shape similar to that of a compound — optionally appear with the izāfa. Crucially, it is the stress bearing property and the syntactic independence that distinguish these non-primary adpositions from the basic ones. Examples of adpositions appearing with the izāfa are *baraw* 'in direction of', *bardaw* 'before, in front of', *labābat* 'concerning', *labāra* 'about', *lagal* 'with', *lalāyn* 'by(passive agent)', *lanāw* 'within', *lapēnaw* 'for the sake of', *lasar* 'on top of, according to', *nāw* 'between, among', *nēwān* 'between, among', *pēš* 'before', *pāš* 'after', *tanīšt* 'next to, beside' (165)-(171).

(165) *labāra-y* *šař-**ī*** *mosil*
about-EZF city-EZF mosul
'about the city of Mosul'

(166) *aw-ān-a-y* *šir-**ī*** *dāyk* *da-xōn-īān* *dwā=y*
that-PL-DEF-EZF milk-EZF mother HAB-eat-3PL PREP=EZF
*la-dayk-bun*
from-mother-born
'those breast fed after being born'

(167) *awa labar-**ī*** *rāsti musel bō merden da-jang-it*
3SG for the sake of-EZF hold Mosul for to die IMPF-fight-2SG
'he/she fights to death to hold Mosul'

(168) *labar* *barzbūnawa **y*** *hal* ***ī*** *bāzargāni*
because of growth EZF opportunity EZF business
'because of business growth opportunities'

(169) *la* *amrīka-wa* *baraw-**ī*** irāq o suryā
PREP America-PREP towards-EZF Iraq and Syria
'from America to Iraq and Syria'

(170) *hīč* *zānyārī-ak-y-ān* *labāra* *anjām-**ī*** *ōparasīōn-aka-wa*
nothing information-DEF-Y-PL about result-EZF operation-DEF-POST
*n-y-a*
NEG-Y-COP
'they don't know anything about the result of the operation'

(171) *awān-īš* *da-čin-a* *nāw* *qutābxāna-kān*
3PL-too HAB-go-POST inside library-DEF.PL
'they go inside the library too'

Like *lagal*, which is often analyzed as an instance of lexicalization, forms ending in -y (*cf.* Table 3.4) such as *dwāy* 'after' (which Thackston (2006) cites as *dwā i* in his Sorani vocabulary) or *bapēy* 'after' can be seen as lexicalization of the collocating prepositions and the izāfa. This can explain why those forms fail to appear with izāfa synchronically. It could also be a phonological constraint given the phonological similarity between the ending of those forms with the izāfa.

To summarize, I have shown in this chapter that the izāfa exhibits properties of a morphological expression that is contextually induced and is sensitive to properties that relate to definiteness. The data suggest that the phenomenon should be thought of in terms of default morphology in dealing with the form alternation and a morphomic distribution given the irregularities observed in the distribution. Before proceeding to a formal proposal in modeling the distribution of izāfa in Sorani, I would like to further deepen some theoretical insights that have emerged from the description I provided here. I thus discuss the question of definiteness, cliticization and the morphological status of the izāfa in the next chapter.

**Chapter 4**

# Theoretical Considerations

In the previous chapter, I have provided a relatively detailed description of the noun phrase in Sorani and have shown that the izāfa marker features an allomorphy sensitive to properties of the noun phrase: The $=\boldsymbol{a}$ form appears within the domain of the marker $=aka$ and its allomorph $\boldsymbol{a}$, which obligatorily appears with the demonstrative $\boldsymbol{am}/\boldsymbol{aw}$ (172)-(173). The $\bar{\boldsymbol{\imath}}/\boldsymbol{y}$ form, on the other hand, appears elsewhere (174)-(175).

(172) *xānū* $\boldsymbol{a}$ *tāz=aka*
house EZF new=DEF.SG
'the new house'

(173) *aw mindāl* $\boldsymbol{a}$ *bčūk-ān-a*
that child EZF little-PL-DEF
'the/those little children'

(174) *řūdāw-ēk* $\bar{\boldsymbol{\imath}}$ *mēžūī*
event-INDF EZF historic
'a historic event'

(175) *kitēb* $\bar{\boldsymbol{\imath}}$ *žīān*
book EZF life
'book of life'

Based on the examination of the data in Chapter 3, I argue that the izāfa is an instance of contextual morphology—that is, its appearance is induced by a syntactic relation between a lexical head and its dependent. But while izāfa marking is systematic, its distribution is not entirely coherent.

1. The izāfa is restricted to head-initial NPs.
2. It appears on nominal and adjectival heads and optionally on non-primary adpositional heads.
3. The izāfa marks a head that has syntactic dependents — specifier, adjuncts or complements.
    - Morphological dependents such as affixes and clitics are not marked with the izāfa.

- The izāfa marks the dependency of nominal complements and adjectival modifiers, with each dependency inducing an izāfa.
- PPs are subject to additional constraints: (1) They optionally induce an izāfa on a head noun marked with *aka* or its allomorph *a*, which appears with the demonstrative. (See section 3.1.3) (2) An adposition may induce an izāfa if it is strictly adjacent to the head. (3) The marking is not recursive as it is with adjectives. Only one PP may induce an izāfa marking.
- CP complements or adjuncts may or may not induce the izāfa. The constraints for CPs are similar to those for PPs in the sense of optionally triggering the izāfa.

Given these observations, I would like to proceed with a brief examination of a few theoretical concepts that I believe are relevant to izāfa constructions and the Sorani NP more generally. I first investigate the issue of definiteness and argue that the allomorphs $=aka/=a$ should be analyzed as expressing specificity rather than definiteness. I then consider the status of clitics and their effect on their hosts. The izāfa as will be shown should be analyzed on a par with markers such as the indefinite, the specific or the plural. They all exhibit properties illustrative of their clitic/affix-like status. Finally, I will argue that the observed distribution of izāfa is morphomic in the sense of Maiden (1992). That is, despite its systemic organization, its distribution is somewhat arbitrary.

## 4.1 Definiteness and Specificity

Classical views on definiteness have used definite descriptions in terms of "referential uniqueness" defining definiteness as "a use of definite description [that] is felicitous if and only if there is exactly one object in the context that satisfies the content of the description" (Zribi-Hertz and Jean-Louis 2013). Löbner (1985; 2011) argues however that definiteness entails non-ambiguity (uniqueness) of identification rather than uniqueness of reference.

> It is not uniqueness [of reference], but non-ambiguity which is essential for definiteness. Non-ambiguity is the property of an expression that allows for only one interpretation (possibly under additional constraints). Uniqueness of reference is always an accidental property

> of a sortal concept... Non-ambiguity, in contrast, may be an inherent property of (also non-sortal) concepts. (Löbner 1985: p.291)

As Löbner notes, the sentence "He is *the son of a famous violinist*" does not identify a unique referent but a unique relational concept for the NP *the son of a famous violinist*. The non-ambiguity of reference necessarily involves a situation or context of utterance i.e. whether non-ambiguity is stated with relation to or independent of the situation. Based on this relation, Löbner also differentiates between semantic and pragmatic definiteness.

> Semantic definites refer unambiguously due to general constraints. Pragmatic definites depend on the particular situation for unambiguous reference. (Löbner 1985: p.299)

Semantic definiteness consists of the non-ambiguity of reference established independently of the situation or the context of the utterance while pragmatic definiteness is the non-ambiguity of reference in relation to the immediate situation or the context of utterance. The basic type of semantic definites indicate *one-place functional concepts* such as *the moon, the sun, the truth* (Löbner 1985; 2011) and they only require a situational argument. Complex types of semantic definites headed by relational nouns (*the son*) involve more than one argument. Proper nouns are also included in the semantic definites (adapted from Zribi-Hertz and Jean-Louis 2013).

### 4.1.1 The marker *=aka* and its allomorph *=a*

The data examined show that the *=a* form of izāfa is sensitive to the marker *=aka* and to the expression of demonstrativeness in the phrase. The marker *=aka* has traditionally been described as a definite marker while the demonstrative marker is analyzed as a circumfix. However, the circumfix account seems to miss a crucial property linked to definiteness. The contrast in (176)-(177) shows that the edge markers *=aka* and *=a* share the same distribution. While *=aka* usually appears without the demonstrative *aw/am*, the latter (as noted in the previous chapter) never appears without the edge marker *=a*.

(176) *šaw-**a*** *sard=aka*
night-EZF cold-DEF
'the cold night'

(Fieldwork data)

(177) *aw* *šaw-**a*** *sard=a*
that night-EZF cold-DEF
'that cold night'

(Fieldwork data)

I propose that these two are allomorphs of the same marker. That is, the $=a$ form of the definite is obligatorily selected with the demonstrative. The fact that both $=aka$ and $=a$ (with *aw/am*) call for the same form of the izāfa (as in (176)-(177)) is evidence of their common property. This is confirmed by the fact that for some speakers, these two forms may alternate, as in (178)-(179).

(178) *aw* *šaw-**a*** *sard=a*
that night-EZF cold-DEF
'that cold night'

(Fieldwork data)

(179) *aw* *šaw-**a*** *sard=aka* *lanāw* *dīsambar*
that night-EZF cold-DEF.SG in December
'that cold night in December'

(Fieldwork data)

### 4.1.2 Weak Definites vs Strong definites (Specific)

It is with Löbner's theory of definiteness in mind that I now reexamine the form alternation of Izāfa described in Chapter 3. A summary of the distribution of the izāfa is provided in Table 4.1.

TABLE 4.1: Izāfa in Sorani Kurdish Noun Phrases

| | Form | DEFINITE SG/PL *-aka/akān* | DEM SG/PL *am/aw...a/āna* | INDF SG/PL *-ēk/-ān* | POSS | BARE |
|---|---|---|---|---|---|---|
| IZĀFA | ***-ī/y*** | * | * | ✓ | ✓ | ✓ |
| | ***-a-*** | ✓ | ✓ | * | * | * |

That the $=a$ form of the izāfa appears with the definite (and demonstrative) follows from my analysis of the demonstrative as expressing definiteness overtly by means of the allomorph of the definite $=a$. However, if it was the property of definiteness that induced the $=a$ form of the izāfa, we would expect to find it

with possessives and genitive constructions such as that in (181) since these often express definite possessive expressions.

(180) *ganj ī am šar a*
young EZF this city DEF
'young (people) of this city'

(McCaurus 2012)

(181) *aw šaw-aka-y kāmyāran*
that night-DEF.SG-EZF kamyārān(name of a town)
'that night of Kamyaran'

(Fieldwork data)

Following Poesio (1994), Barker (2004) explains that there is a productive, systematic class of definite descriptions whose usage does not need familiarity or uniqueness; these are called WEAK DEFINITES. For example, in the sentence "I hope the cafe is located on **the corner of a busy intersection**." (Poesio 1994), neither the speaker nor the listener has familiarity with the a specific intersection or corner, nor is there an implication that the intersection in question has only one corner. Barker (2004) claims that weak definites require the presence of a relational head noun and an overt genitive prepositional phrase argument, e.g. *corner* is a syntactically and semantically relational and *of a busy intersection* is a genitive *of*-phrase in English. Weak definites are identified by the degree of uniqueness in various cases. While a *definite description* is an NP with non-trivial descriptive content that is determined by *the* in English, a *possessive definite* is "one whose head nominal occurs in a construction with a genitive *of* prepositional phrase argument" (*the corner of a busy intersection*) (Barker 2004). For Barker, possessive definite descriptions may be used in contexts in which more than one object is satisfying the content of the description. This quality of a definite description in such a context makes it a weak definite.

Sorani possessives behave similarly, appearing with genitive prepositional phrase arguments. Since this is yet another type of nominal dependency, an izāfa appears between the possessor and the possessum. However, in possessive constructions, it only appears as *ī/y* rather than *a*, which is surprising if we assumed that the latter marked definite nouns. However, one might argue that the type of definiteness expressed in possessives corresponds to Barker (2004)'s weak definites or Löbner (1985; 2011)'s semantic definite compared to nouns specified by *=aka*

and its allomorph which would better be described as pragmatic definites. Possessive phrases usually express semantic definiteness in providing non-ambiguity of reference established independently of the situation or the context of the utterance. They are weak definites only requiring the non-familiarity with a specific reference in Barker (2004)'s sense, identified by a degree of uniqueness.

(182) *nāw ī mudīr ī ma'ārif*
name EZF director EZF education
'the name of the director of education'

(McCaurus 2012)

(183) *bās ī mēžū ī kurdustān*
discussion EZF history EZF Kurdistan
'discussion of the history of Kurdistan'

(McCaurus 2012)

I propose that the reason we do not see the same form of the izāfa appearing with possessive is precisely because they are weak definites compared to *=aka* or the demonstrative which I assume are strong definites. The implication is that only strong definites induce the the *=a* form of the izāfa. This means that definite readings of bare plurals, denoting for instance, kind or generic readings, would likewise not exhibit the *=a* form of the izāfa such as *dodo* in (184) and *beach* in (185).

(184) *min bās ī dōdō da-ka=m*
1SG describe EZF dodo IPFV-do=1SG
'I talk about the dodo'

(Glosbe 2018)

(185) *lasar řōx pālkawtin* o *xō dāna bar xōr*
on beach lie down and REFL give in front of sun
'lie on the beach sunbathing'

(Glosbe 2018)

The properties of the definite *=aka/=a* and the demonstrative *aw/am* are in fact consistent with Löbner (1985; 2011)'s pragmatic definites or what has usually been identified as specific, i.e., they involve non-ambiguity of reference in a specific situation. Hence, it would seem that the conditioning environment in which

the allomorphy occurs is not simply definiteness but more precisely that of pragmatic definiteness or specificity. The denotation of specificity by demonstratives has been noted to be a property available cross-linguistically. Similarly, there are languages, that are typologically different from Kurdish but in which the demonstrative marker must, as in Kurdish, appear with the specific marker. For example, French-based creoles like Mauritian, Guadeloupean or Haitian precisely have NP constructions where demonstratives always require the presence of a specificity marker as in (186).

(186) Sa zanfan *(la)
DEM child SPC
'This child' (Mauritian Creole, p.c. Fabiola Henri)

### 4.2 Clitic or affix?

In this section, I investigate in more detail the status of markers like the specificity marker ***aka/a***, the indefinite marker ***ēk***, the demonstrative ***am/aw... a***, the plural marker ***ān*** and some simple prepositions such as *ba*, *la*, *bē*, *bō* and, centrally, the izāfa marker.

I first consider the indefinite marker ***ēk*** and the demonstrative ***aw/am***. Both of these markers show strict attachment to noun lexemes: ***ēk*** is found postnominally (as in (187)) while ***am/aw*** precede the noun *modulo* the constraint for the specificity marker to appear(as in (188)). Hence, the indefinite marker and the demonstrative have the affixal property of selecting a particular category of host.

(187) *dāristān=**ēk** ī gawra*
forest-INDF EZF large
'a large forest'

(188) ***aw**=mindāɬ=a bčūk=ān=**a***
that=child-EZF little=PL-DEF
'those/the little children'

The fact that no element can intervene between the head noun and the demonstrative marker would support its analysis as an affix. However, evidence from coordination suggests that these markers do not strictly appear on the head noun but can scope over a coordination of lexical nouns or modified nouns (as in (189)-(192)) them perfect candidates for a clitic analysis.

(189) ***aw**=[pyāw=a barz=ān=**a** o āfrat=a jwān=ān=**a**]*
that=man=EZF tall=PL=DEF and woman=EZF beautiful-PL=DEF

'Those tall men and beautiful women'

(Fieldwork data)

(190) ***aw***=[*pyāw=a barz=ān=**a*** o ***aw***=*āfrat=a jwān=ān*]=***a***
that=man=EZF tall-PL=DEF and that=woman=EZF beautiful-PL=DEF
'Those tall men and beautiful women'

(Fieldwork data)

(191) **aw=pyāw=a barz* o *āfrat=a jwān-ān=a*
that=man=EZF tall and woman=EZF beautiful-PL=DEF

(Fieldwork data)

(192) **aw=pyāw=a barz=ān=a* o *āfrat=a jwān*
that=man=EZF tall=PL=DEF and woman=EZF beautiful

(Fieldwork data)

As these examples show, the demonstrative *aw* can either appear once on a coordination of nouns modified by adjectives (189), or can be marked on each conjunct (190). Yet as (189) shows, the specificity marker *a* that obligatorily appears with the demonstrative needs to be marked on both conjuncts. Similarly, the indefinite is an enclitic to nouns only but can also mark a coordination of lexical nouns as in (193)-(196).

(193) *pyāw=**ēk** o āfrat=**ēk***
man=INDF and woman=INDF
'a man and a woman'

(Fieldwork data)

(194) *pyāw o āfrat=**ēk***
man and woman-INDF
'a man and woman'

(Fieldwork data)

(195) **pyāw=**ēk** o āfrat*
man=INDF and woman
'a man and woman'

(Fieldwork data)

(196) *pyāw=ī barz o āfrat=**ēk***
man=EZF tall and woman-INDF
'a tall man and woman'

(Fieldwork data)

The scopal effects and marking of a phrasal (or complex) construction observed above is what is expected from clitics compared to affixes. Notice that according to my informants, there is a preference for repetition of the markers in general coordination.

The plural marker *ān* shows peculiar properties. First, it exhibits a case of variable ordering that seems to be sensitive to the phonological environment. It follows the specificity marker *aka*, as in (203), but it precedes the specificity

marker's allomorph *a* (which accompanies the demonstrative), as in (204). Then, we may observe that the plural cliticizes to different hosts in the presence/absence of other markers: It is enclitic to a bare noun (197) and expresses an indefinite plural or a bare plural. It is also postposed to the specific *aka* (203) and (208). Finally, as with previous markers, the plural can scope over a coordination of bare nouns as in (201).

(197) *pyāw=**ān**=ī barz*
man=PL=EZF tall
'tall man'

(Fieldwork data)

(198) *?pyāw o āfrat=**ān***
man and woman=PL
Int. 'men and women'

(Fieldwork data)

(199) **pyāw=**ān**=ī barz o āfrat*
man=PL=EZF tall and woman
Int. 'tall men and women'

(Fieldwork data)

(200) **pyāw=ī barz o āfrat=**ān***
man=EZF tall and woman=PL
Int. 'tall men and women'

(Fieldwork data)

(201) *pyāw o āfrat=ān=**ī** barz*
man and woman=PL=EZF tall
'tall men and women'

(Fieldwork data)

However, scope over a coordination of phrasal construction is more constrained. Consider again the examples with the demonstrative in (189)-(190) above. The plural is repeated (with the specific) on both conjuncts despite the fact that the demonstrative marks both of them. Notice also that in these examples, the plural marker can be said to be unselective, appearing in these examples they appear on the adjective in NP-final position of the phrase rather than on the head noun.

The specificity marker ***aka*** and its allomorph ***a*** (202)-(207) occur at the right edge of the nominal phrases. Like the plural, the specificity marker is not selective of its host, which can be a noun or an adjective and they move to the right edge of the phrase regardless of the multiple number of modifiers after the head noun in the phrase. They are in a way similar to the English possessive 's with the difference that phrases like PP or CPs fall outside of the nominal domain (cf. Samvelian 2007a). Like the indefinite and the demonstrative, the specificity marker can also scope over coordination of lexical nouns. But again, speakers show a preference for repetition of the marker.

(202) *kitēb=**aka***
book=DEF.SG
'the book'

(203) *xāł-a sūr=**ak=ān***
dot-EZF red=DEF=PL.
'the red dots'

(Fieldwork data) (Glosbe 2018)

(204) *am/aw kitēb=ān=**a***
this/that book=PL=DEF
'this/that book'

(205) *aw šaw-a sard-**a***
that night-EZF cold-DEF
'that cold night'

(Fieldwork data) (Fieldwork data)

(206) *kitēb=**aka** ba zmān ī farsī*
book=DEF.SG in language EZF Persian
'the book in Persian'

(207) **kitēb ba zmān **ī** farsī=**aka***
book in language EZF Persian=DEF.SG
'the book in Persian'

(208) *pyāw o āfrat=**akān***
man and woman=DEF.PL
'The men and women'

(209) **pyāw=**akān** o āfrat*
man=DEF.PL and woman
'the men and women'

(Fieldwork data) (Fieldwork data)

Finally, as concerns the izāfa, the fact that it features an allomorphic variation that (i) morphosyntactically distinguishes between specific and non-specific and (ii) is phonologically conditioned (*ī* vs *y*) favors a morphological analysis. The non-selectivity of its host despite the strict attachment favors a clitic approach rather than an affixal one: they mark non-primary prepositions, nouns and adjectives and like other markers discussed previously in this chapter, the izāfa can mark a coordinated construction rather than attaching to every lexical head that construction.

(210) *pyāw o žin **ī** bāłābarz*
man and woman EZF tall
'tall man and woman'

(211) *pyāw **ī** barz **ī** jwān*
man EZF tall EZF handsome
'tall handsome man'

(Fieldwork data) (Fieldwork data)

This is not necessarily an instance of scope as we saw with the other markers. Rather, as we mentioned before, the izāfa is contextual morphology and surfaces

as a result of a syntactic dependency. The data favors an approach in terms of constructional locality where "constructions licence mother-daughter configurations without reference to embedding or embedded contexts" (Sag 2012: p. 150). Constructional Localism requires all dependencies, local or non-local, be locally encoded in signs in such a way that "information about a distal element can be accessed locally at a higher level of structure."(Sag 2012) It is precisely in those relations, albeit with the constraints described in Chapter 3 that the izāfa appear (Figure 4.1).

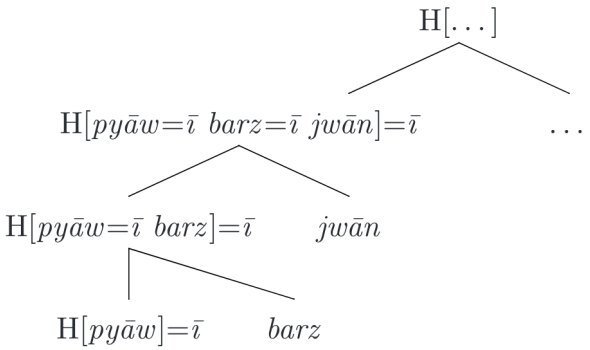


FIGURE 4.1: Izāfa in constructional locality

As noted, its distribution in the structure is incoherent, that is, not all syntactic dependencies induce the izāfa. I turn to this point in the next section.

## 4.3 Morphomic distribution

The term morphome was originally coined by Aronoff (1994) to "name a level of representation between morphosyntactic feature arrays and their phonological realization, in particular to single out cases where more than one morphosyntactic feature array maps onto the same set of phonological realizations" (Aronoff 2016). Morphome refers to patterns of morphological realization that are not motivated by phonology, syntax or semantics but rather involve a systematic mapping between arbitrary classes of morphosyntactic features and arbitrary sets of morphophonological forms (Aronoff 1994: 28). There are cells in the paradigms of languages that have the same inflectional forms but have distinct contents in a way that they do not form a natural class. In such cases, Aronoff (1994) believes that the inflectional morphology of the language refers to categories that are irrelevant for

the semantic interpretation and syntactic distribution of words in syntax embodying a mismatch between form and function (adapted from Taghipour (2017)). In more recent morphological research, morphomes are sometimes seen as patterns in paradigms that are not syntactically or semantically defined but which morphology reinforces. Aronoff's original example for the existence of morphomes is the English perfect/passive participle in which the passive or the perfect map onto the morphological function $F_{en}$ which realizes the passive/perfect form depending on the verb wohse participle is going to be realized. The functions of this type that map one of many morphosyntactic feature arrays onto one of many realizations are morphomes (Aronoff 2016). "Morphomes are functions within an [inferential]-realizational theory of morphology that map morphosyntactic representations onto phonological realizations. Their job is to link levels of representation and so they cannot be natural, by definition." (Aronoff 2016). The fact that whether the morphome realizes an array of morphosyntactic features or a disjunction refers to monovalent or polyvalent syntax and whether the morphome has one realization or the realization is context sensitive refers to the monomorphous vs polymorphous identity of the morphome. Based on the number of distinct feature arrays and the number of distinct realizations, Aronoff (2016) proposes four specific types of morphomes (Figure 4.2). First, **monovalent monomophous** such as the accusative suffix *-annu* in Kannada which marks all accusatives including pronouns. Second, **polyvalent monomorphous** such as English *-Z* which encodes plural in nouns, third-singular present indicative in verbs, possessive and contractions (those of *is*, *has*, and *does*). Third, **monovalent polymorphous** such as English past tense that has variety of realizations depending on the individual verb. Fourth, **polyvalent polymorphous** such as English $F_{en}$ whihc has two morphosyntactic values and maps onto many realizations depending on the individual verb (Aronoff 2016).

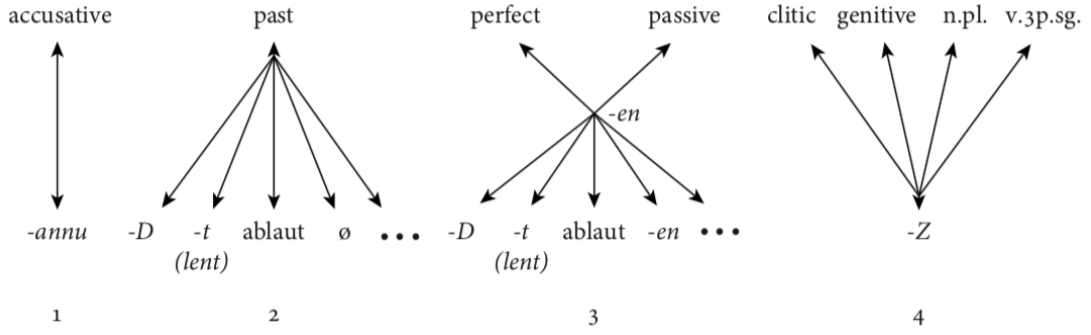


FIGURE 4.2: Graphic representations of the four types of morphomes (Aronoff 2016)

As Bermúdez-Otero and Luís (2016) put it, from this perspective, morphology is seen as specifying functions from syntactic feature arrays to phonological

realizations. In a narrow sense, they consider the existence of polyvalent polymorphous patterns to motivate a "hidden morphological level of representation: the morphomic level." (Figure 4.3)

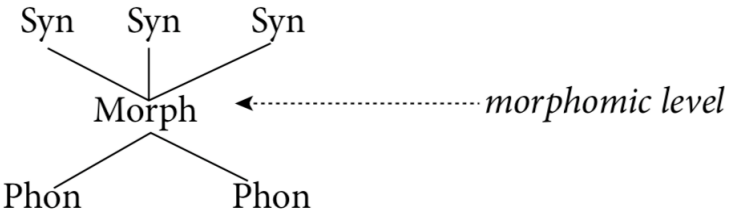


FIGURE 4.3: Morphomic level (Bermúdez-Otero and Luís 2016)

The notion of the morphome takes a dimension in the work of Maiden (1992), which examines the distribution of morphomic patterns that survive through the evolution of Romance languages. His idea is that languages exhibit patterns of alternation that are phonologically and functionally independent but whose distribution is systemic. That is, while distribution doesn't seem to have stable phonological or functional motivation and is arguably 'irregular', speakers can predictively use the system.

In chapter 2, I briefly mentioned a historical account of izāfa in Northern Kurdish and Persian by Haig (2011). Haig claims that the izāfa in Old Persian had two functions: Relative Pronoun and Demonstrative/Anaphoric. He states that in Northern Kurdish, izāfa has retained these functions, while in modern Persian, it has lost these functions and izāfa has changed to an adnominal Persian-style linker. Haig (2011) suggests that the history of the izāfa in West Iranian is a good illustration of the notion of exaptation. The term 'exaptation' is first used by Lass (1990) as a metaphor from evolution in biology. Exaptation in language is used to describe instances of morphology which have been productive in the past but through changes elsewhere in the system become functionally redundant. Productive morphology, especially if it is frequent, will not generally disappear but may instead attract a new function (Haig 2011). The changes to the izāfa in Persian are analyzed with this approach by Haig (2011). He believes that with the loss of case and gender and the intrusion of a complementizer, the izāfa has lost its relativizing, anaphoric and complementizing functions yet has retained its presence in the NP.

The function of the izāfa in Sorani Kurdish is in line with this approach. That is, while the form alternation is morphosyntactically transparent or straightforward, its distribution, despite being induced by syntactic constraints, is less

so and may be characterized as morphomic. In other words, izāfa has a morphomic distribution. Based on the discussion in §4.1 and §4.1.2, the choice of the izāfa form *ī* over *a* is not difficult since the contrast lies in the distinction between specific and non-specific with the morphomic function $F_{EZF}$ that maps the morphosyntactic arrays into the phonological forms of izāfa (***ī/y***, ***-a-***), yet, its distribution is incoherent and not quite predictable with the appearance of syntactic categories like PPs and CPs. Such exaptation is expected if we acknowledge the predictive nature of the pattern without committing to the strict isomorphism that structuralist approaches tends to assume. In other words, form and function might change over time but the morphomic or predictive nature of the izafa persist throughout language change.

To summarize, in this chapter I have shown that the form of izāfa is sensitive to the multiple levels of definiteness in a phrase in Kurdish. Drawing on the previous studies on definiteness and specificity, weak and strong definites, izāfa represents two allomorphic appearances corresponding with the specificity of the phrase. In the second section, I examine the distribution of the definite markers *aka/a*, indefinite *ēk*, plural marker *ān* and the izāfa *ī/y* or *a* in order to investigate the clitic- or affix-like behavior of these markers. While the markers show various levels of selectiveness, there is evidence for their clitic-like behavior. In the last section, drawing on the notion of exaptation and a historical account of izāfa in Northern Kurdish and Persian, I associate the idea of morphome with the distribution of izāfa due to the predictable but incoherent appearance of izāfa.

**Chapter 5**

# Avenues for a Constructional Approach to the Izāfa Construction

## 5.1 Construction Grammars: HPSG and SBCG

In returning to the traditional 'taxonomic' type of grammatical analysis, the advocates of CxG (Construction Grammar) have suggested that "there are constraints on form and interpretation that cannot be explained as the product of grammatical constructions, form-meaning pairings of varying degrees of productivity and internal complexity" (Michaelis 2006). In CxG, grammar is represented as a 'structured inventory' of the pairings of form and meaning in a language. In his account of the advantages of construction grammar, Sag (2012) mentions its direct concern with words, generalizations about lexical classes and the patterns according to which complex expressions are constructed. In short, the underlying assumptions for BCG(Berkley Construction Grammar) and HPSG(Head-Driven Phrase Structure Grammar) are as follows:

- Linguistic objects are modeled in terms of feature structures that are represented as attribute value matrices (AVMs) or directed graphs.
- Feature values can be complex, even recursive.
- A language consists of a set of signs; a sign is an abstract entity that is the point of constraints on the interface of form and meaning.
- A grammar is a system of constraints that work together to license and delimit the signs of a given language.

- *Constructions*, the constraints on classes of signs and their components, are organized into a regime, a lattice-like array of types and subtypes that allows generalizations of various degrees of granularity to be stated simply.
- The distinction between lexical and grammatical entities is blurry, motivating a uniform conception of lexical and constructional constraints.

(adapted from (Sag 2012))

HPSG is a constraint-based theory initially modeled after GPSG and developed by Pollard and Sag (1994). A constructional version of the model is proposed in Sag et al. (2003) and Boas and Sag (2012). HPSG is a generative theory of grammar which shares with other theories of grammar the goal of modeling how human language is structured in the human mind. HPSG uses a rich lexicon and the Attribute Value Matrix (AVM) to represent lexical information. The main tool in HPSG is the features that are used to show the category of a word, the arguments that a word might take or its role in the construction of a sentence.

HPSG is strictly lexicalist—that is, the lexicon is not just an encyclopedia of words, it is deeply structured. In HPSG, parts of speech, lexemes, syntactic, semantic and valence rules in the language are organized into types based on shared properties. The organization of such types is considered as a hierarchy. The constraints and features belonging to each type are inherited by their subtypes to the individual types at the very bottom of the hierarchy and the inheritance of the constraints in the hierarchy of types gives rise to a formulation of defeasible constraints. Natural languages do show regularities inherited by the subtypes but there are always exceptions to the default rule. This leads to using another understanding of the constraints on the hierarchy, one that is defeasible. Defeasible constraints hold by default unless being overridden by some other constraint at a more specific level. In other words, the contradictory information related to the subtype takes precedence over the default constraint. In other words, the inheritance hierarchy is nonmonotonic. By organizing the the lexicon as a type hierarchy with the use of default constraint inheritance, the stipulations related to the particular lexical entries are reduced and shared properties of the different word classes are expressed in a more general way. In addition, any exceptions to the default constraint are permitted. Each lexical entry in a lexical hierarchy includes a distinct family of lexemes that are all types of a supertype. The members of the family inherit the constraints in the given lexical entry and those of the supertypes. A lexeme inherits all the default and the defeasible constraints. The lexical entries

with the constraints inherited through the lexeme hierarchy form the set of basic elements of the language giving rise to a family of lexical sequences which include an input and an output, the latter being a feature structure of the type word (adapted from Sag et al. 2003).

In SBCG (Sign Based Construction Grammar), language is considered as consisting of an unlimited number of the notion of Saussurean 'signs' but encompassing more components, i.e., phonological structure, (morphological) form, syntactic category (word's syntactic category and combinatoric potential), semantics (the frames that collectively define the meaning of a word, a word's referential index), and contextual factors including information structure (Michaelis 2009, Sag 2012). Sign-Based Construction Grammar assumes that "grammar is an inventory of signs—complexes of linguistic information that contain constraints on form, meaning and use—and constructions are the means by which simpler signs are combined into more complex signs"(Michaelis 2009). SBCG models signs as *feature structures*. The values of features can include word classes such as *adjective*, case values such as *genitive*, referential indices, lexeme classes such as *ditransitive-verb-lexeme*, the forms of words such as *pick* and the binary values + and –. Grammatical categories are modeled through the complexes of properties; for example, nouns have values specified values for CASE, NUMBER and GENDER and verbs have values specifie such as [VFORM *finite*] or [VFORM *present-participle*]. Feature structure are recursive; that is, the value of a feature may be another feature structure. As construction grammars, both HPSG and SBCG are constraint-based models which do not assume movements, underlying structure or empty categories. SBCG uses semantic constraints and phrase structure rules to define constructions. It is notable that in SBCG, "phrase-structure rules mean things in the same way that words do" (Michaelis 2009). The phrase-structure rules of SBCG combine constraints on meaning, use and form.

In Sign-Based Construction Grammar (SBCG) the most important model objects in the language are signs and the formal representations of signs are actual words, phrases and sentences. Each lexical sign or fixed phrasal expression is licensed by a **listeme**, a 'listed' description of a word or phrase. The signs of SBCG are not just words and lexemes but they also include phrases and sentences, i.e., the FORM value of a phrasal sign is a list of words. Expressions can either be covert, that is, gapped or involving a silent pro for those languages that would need such stipulations while overt expressions have subtypes, phrases and words (Figure 5.1). Signs and constructs are modeled by feature structures in SBCG. In formal terms, a construct(*cxt*) is a feature structures that includes a mother(MTR)

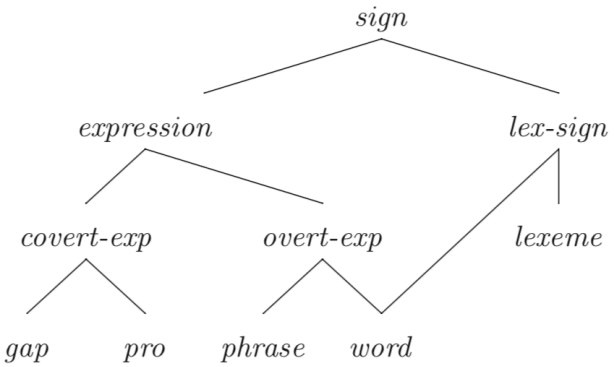


FIGURE 5.1: Hierarchy of signs, expressions, lexemes and words (Sag 2012).

and a daughter(DTR) feature. A construct always encompasses of a mother sign and a non-empty list of daughter signs. Signs modeled in Feature Structures (FSs) are of the types *atoms* and *functions*. Features can be subsets of the set of feature. A **functional** FS maps a feature in its domain to an appropriate value (atom or value), i.e., functional FSs map features to feature structures. A unique feature of SBCG is the construct, local trees that are licensed by a particular kind of construction called a combinatoric construction. Constructions are formalizations in a constraint-based system.

## 5.2 Some hints for a constructional approach using SBCG

SBCG uses combinatory constructions to account for phrase types in a language. Combinatory constructions describe constructs of a language and constructs are made of one or more distinct signs. In other words, "constructs are local trees (mother-daughter configurations) with feature structures (specifically, signs) at the nodes. Constructions can describe only such mother-daughter dependencies and not, for example, mother-granddaughter dependencies" (Michaelis 2009, Sag 2012). SBCG uses local trees that qualify as constructs in modeling a language rather than derivational trees. As mentioned earlier, a construct is modeled by a feature structure in SBCG that contains a mother (MTR) and daughters (DTRS). The value of the MTR feature is a sign and the value of the DTRS feature is a list of one or more signs (Michaelis 2009).

In chapter 4, I mentioned that the surfacing of the izāfa is an instance of syntactic dependency and its form is dependent on contextual morphology. The

evidence from the clitic-like behavior of izāfa shows that it attaches to lexical entries and structures bigger than a word, i.e., izāfa targets a structure that is a construct. Based on what we saw in Figure 4.1 and the figure below, we know that izāfa is postposed to lexical constructs or phrasal constructs. To account for the distribution of izāfa in Kurdish phrases, I follow Sag (2012)'s definition of constructional localism in which it is required that all (local and non-local) dependencies are locally encoded to a sign in the way information about a distal element can be accessed locally at a higher level of structure.

(212) *pyāw* **ī** *barz* **ī** *jwān*
man EZF tall EZF handsome
'tall handsome man'

(Fieldwork data)

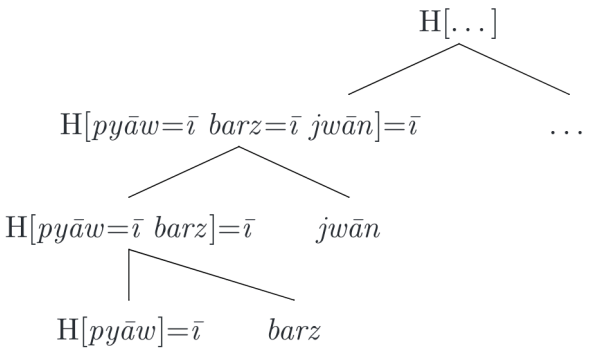


FIGURE 5.2: Izāfa in constructional locality

This analysis is couched within sign-based construction grammar. The iterative nature of the izāfa discussed in the previous chapters, its non-selectiveness and the hosts chosen by izāfa, presuppose that this particle attaches to constructs that either select an argument or are selected by a functor. In addition, given the structure in Figure 5.3, it is also suggested that the attachment of this marker can target either lexical constructs for nouns with a single dependent or phrasal constructs for heads that have already undergone modification. Thus, the hierarchy in Figure 5.3 allows for *phrasal-cxt* to share with *lexical-cxt* the subtype *postinflectional-cxt*. An *izāfa-cxt* is introduced which inherits properties from the *post-inflectional-cxt*. This construction takes as its daughter either a lexical or a phrasal construct and applies the *izāfa* function $F_{EZF}$ in order for the mother node to be marked by the default *ī/y* form of izāfa.

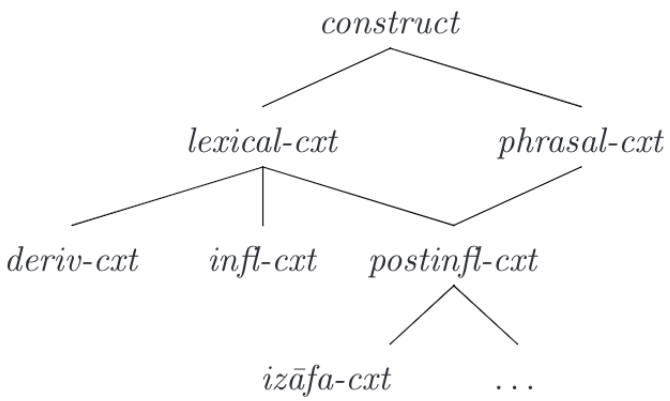


FIGURE 5.3: Phrasal hierarchy in Kurdish

To account for the categories that inflect for izāfa when they have syntactic dependencies, nouns, adjectives and non-primary prepositions are considered as nominals in the hierarchy. This allows the non-primary prepositions to share with nouns and adjectives the property of hosting the izāfa.

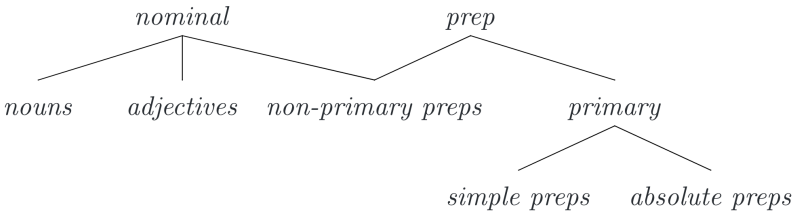


FIGURE 5.4: Nominal hierarchy in Kurdish

I follow Samvelian (2007b) in using the feature DEPS ± to encode complement, specifier and adjunct dependencies. This boolean feature allows izāfa to be specified on nominals since they are the only categories that get head-marked. A construct including the lexical entries and phrases will be marked as DEPS + if their feature is a non-empty valence list or if it selects modifiers. First, we need a rule that induces the izāfa as a type of head marking morphology with a function that is sensitive to the appearance of the dependents of the constructs. As stated

in table (5.1), izāfa suffixation rule suffixes the $\bar{\imath}/y$ form of the izāfa to a construct as the default suffixation. This rule includes all the instances of the appearance including the possessive, the indefinite, the bare nouns and the prepositions.

PPs as dependents trigger the marking when they appear FIRST on the ARGS list. If the PP appears on the REST ARGS list, we allow for an outscoped construction to move it to the EXTRA list next to CPs if present. CP dependents always appear on the EXTRA list since as we argued they are always extraposed.

Izāfa Suffixation Rule

(213)
$$\begin{bmatrix} \textit{pl-nom-cxt} \\ \text{PHON } \boxed{1} \\ \text{CAT } \begin{bmatrix} \text{DEP} & - \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{izāfa-cxt} \\ \text{PHON } \text{F}_{\text{EZF}} \ (\boxed{1}) \\ \text{CAT } \begin{bmatrix} \text{DEP} & + \end{bmatrix} \end{bmatrix}$$

In addition, since the izāfa exhibits the *a* form within the scope of the specificity marker *aka* and the demonstrative marker *am/aw...a*, we need a specific rule that changes the marking of the izāfa. Once the value of the feature DEF is specific, the izāfa changes the form from $\bar{\imath}/y$ to *a*.

Specific Rule

(214)
$$\begin{bmatrix} \text{INPUT} \left\langle \text{MRKG} \mid \text{EZF} + \right\rangle \\ \text{OUTPUT} \left\langle \text{MRKG} \begin{bmatrix} \text{EZF} + \\ \text{DEF} \ \textit{spf} \end{bmatrix} \right\rangle \end{bmatrix}$$

The izāfa function uses a default and a specific form to attach the izāfa to the constructs (Table 5.1).

TABLE 5.1: Izāfa morphological function

| X | $F_{EZF}(X)$ |
|---|---|
| Default | X-*ī/y* |
| Specific | X-*a* |

Figure 5.5 is a tree exemplifying the postnominal adjective sequences which exhibit the recursive use of izāfa. Once adjectival dependents appear after a head, izāfa is induced recursively. Figure 5.6 shows a postnominal adjective and PP. Izāfa exhibits no recursion because of an adjacency requirement of the PP. The non-adjacency of the PP is illustrated in the EXTRA list to show that the PP is out of the scope of izāfa and hence izāfa is not induced.

(215) *pyāw* ***ī*** *barz* ***ī*** *jwān*
man EZF tall EZF handsome
'tall handsome man'

(Fieldwork data)

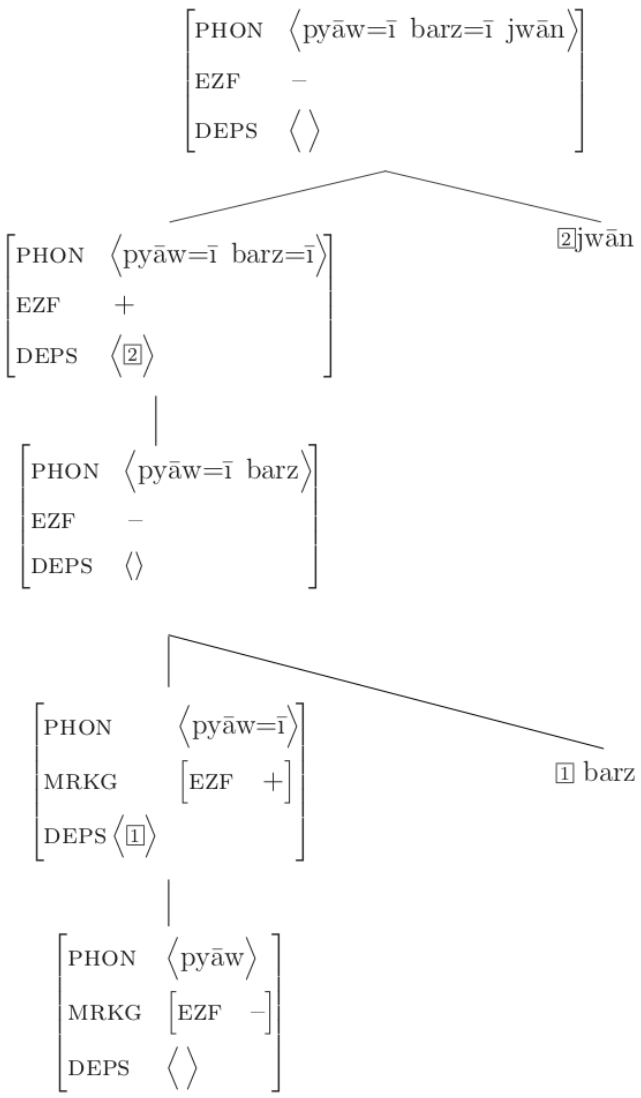


FIGURE 5.5: Izāfa occurence in postnominal adjective sequences

(216) *kitēb=ēk* ***ī*** *kūrdī* *lasar Linux Ubuntu*
book=INDF EZF Kurdish PREP Linux Ubuntu
'a Kurdish book on Linux Ubuntu'

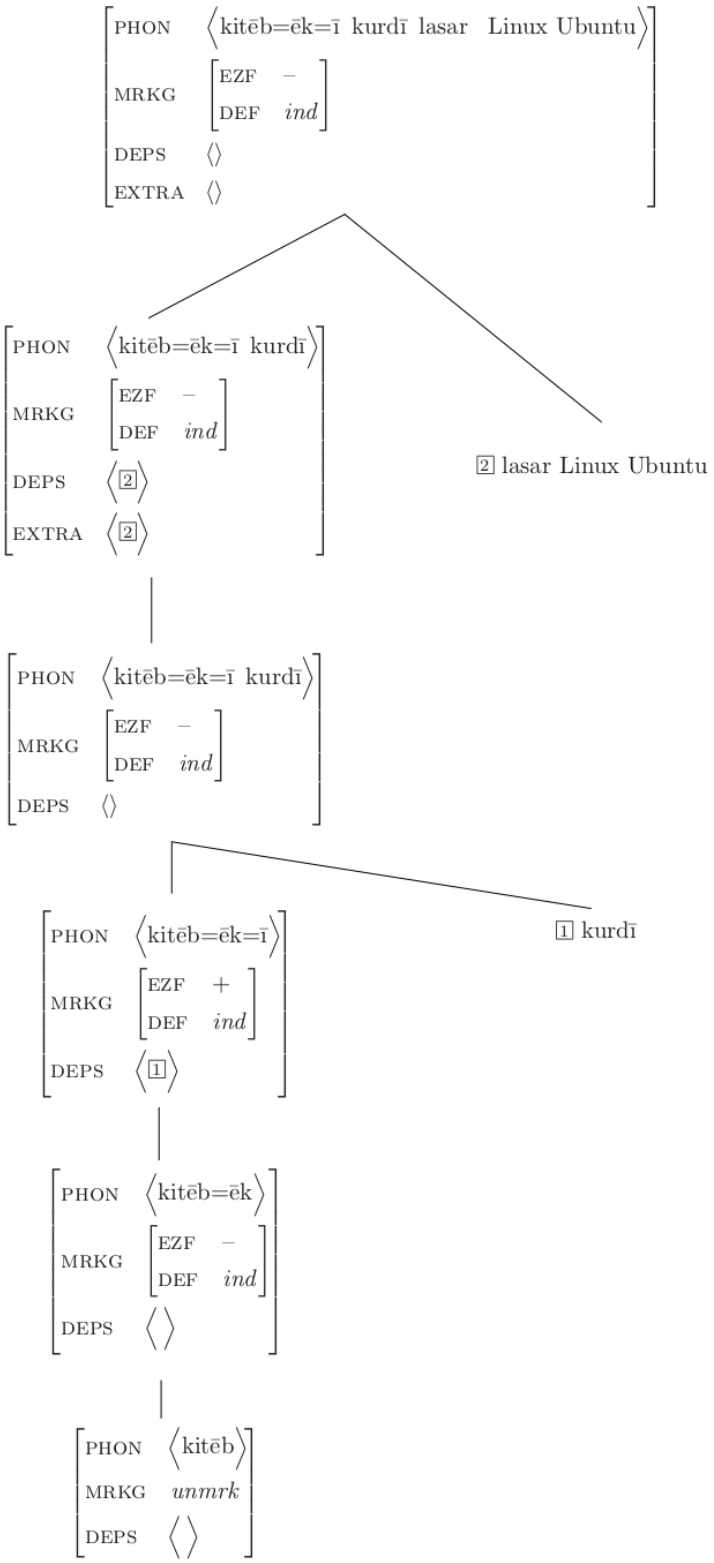


FIGURE 5.6: Izāfa occurence with a postnominal adjective and a PP

This chapter provided some hints to account for izāfa in construction-based grammar. It suggests that since izāfa is attached to lexical items or phrases, we can account for it in SBCG by considering it as a post-inflection to a construct. Nouns, adjectives and non-primary prepositions are considered as nominals that host the izāfa in Kurdish. Given that, we have a suffixation rule with a function for izāfa that attaches izāfa to constructs, functors or complements, and another rule that overrides the default form of the izāfa to account for its form in the specific constructs.

# Chapter 6

# Conclusion

## 6.1 Summary and Conclusion

I have examined in this research the izāfa construction in Sorani Kurdish. Izāfa is a common particle across many Iranian languages representing different different functions and forms. The study is a relatively detailed analysis of the distribution and form alternation of the izāfa in various NP constructions of Sorani Kurdish. The data is predominantly taken from the corpus of Sorani Kurdish along with some elicitations and grammaticality judgments of native speakers.

I have analyzed different inflections on nominals of Sorani Kurdish including nouns, adjectives and non-primary prepositions. In Sorani, the izāfa marker features an allomorphy sensitive to properties of the noun phrase: The $=\boldsymbol{a}$ form appears within the domain of the marker $=aka$ and its allomorph $\boldsymbol{a}$, which obligatorily appears with the demonstrative $\boldsymbol{am}/\boldsymbol{aw}$. The $\bar{\boldsymbol{\imath}}/\boldsymbol{y}$ form of izāfa, on the other hand, appears elsewhere. Since the appearance of izāfa is induced by a syntactic relation between a lexical head and its dependent, I argue that the izāfa is an instance of contextual morphology. But while izāfa marking is systematic, its distribution is not entirely coherent. There are constraints on the appearance of izāfa. It always appears on head-initial NPs attaching to nouns, adjectives and optionally to non-primary adpositional heads. In fact, izāfa marks a head selecting syntactic dependents like specifiers, adjuncts or complements. This means that morphological affixes or clitics are not marked by izāfa. PP and CP complements and adjuncts are subject to more constraints since they optionally induce the izāfa. The PPs strictly require adjacency to the head to trigger the izāfa.

Izāfa is sensitive to the multiple levels of definiteness in Kurdish. Drawing on the previous studies on definiteness and specificity and on weak and strong

definites, it is shown that izāfa represents two allomorphic appearances with sensitivity to specificity of the noun. I have examined various distributions of the definite markers *aka*/*a*, indefinite *ēk*, the plural marker *ān* and the izāfa *ī*/*y* or *a* to investigate the clitic or affix like behavior of these markers. While the markers show various levels of selectiveness, there is evidence that they are clitic-like in their behavior. The distribution of izāfa is argued to be morphomic in Kurdish, both because the conditions on its appearance, though predictable, are incoherent and because its appearance in NPs is sometimes optional.

A short construction-based grammar approach is taken to account for the incidence of izāfa and for its alernation in form. This approach suggests that since izāfa is attached to lexical items or phrases, we can account for it in SBCG by considering it as a post inflection to a construct. Nouns, adjectives and non-primary prepositions are considered as nominals that host the izāfa in the Kurdish nominal hierarchy. Given that, we have a suffixation rule with a function for izāfa that attaches izāfa to constructs, functors or complements, and another rule that overrides the default form of the izāfa to account for its form in the specific constructs.

## 6.2 Recommendations for further research

During the course of the study, some questions arose that require further investigation. I have listed the questions with some limitations of the study below.

- Sorani Kurdish uses a Perso-Arabic alphabet for its script. To investigate the corpus and find the instances in which a preposition appears with or without izāfa, I have used R to analyze the corpus. Sorani Kurdish uses a number of combined prepositions that begin with *la* or *ba*. An adposition like *lasar* contains *la* and *sar* separated by a half space[1] in written text of Kurdish. In order to search the corpus of the present work, I have copied the combinations with half space from the corpus in R. To deal with half space, a better solution for refining the analysis and getting better data would be to use regular expressions (regex) in searching the corpus. For example, the 'negative look ahead' feature in regular expressions would allow one to

[1] A half space or FOUR-PER-EM SPACE (mid space) is a zero-width non-joiner (ZWNJ), a non-printing character used in the computerization of writing systems that make use of ligatures. When a half space is used between two characters, the characters are printed in their final initial forms otherwise, they will be connected into a ligature. The half space is encoded in Unicode as U+200C. See : https://en.wikipedia.org/wiki/Zero-width$_n$on-joiner (Last checked: 07-23-2018)

identify the instances of *la* that are not followed by any other words to make a combined adposition. In addition, the texts in Sorani are not consistent in their use of Unicode, specifically with regard to the letter *a*. *la* may appear in a text either with the isolated form of *a* or the final form of *a* in Sorani Kurdish. In the present work, I searched for both the isolated and the final form of *a* to look for any word that included it. This issue can be approached by using alternation (|) in regular expressions.

- There is not a part-of-speech (POS) tagger or a lemmatizer for Sorani Kurdish. Part-of-speech tagging has been done manually only in one corpus, Kurdish Language Corpus, in which lemmas are not considered in tagging. That is, *kitēb* 'book' and *kitēb=aka* 'the book' are two unique types, rather than tokens of the same type. Providing a part-of-speech tagger that can identify clitics such as *ēk*, *aka*, *akān* and *ān* in Kurdish would be of great help for the analysis developed in the present work. For example, finding the instances in which *aka* is followed by izāfa before CPs and PPs would shed light on the constraints on for the optionality of the triggering of izāfa in these cases.

- In some dialects of Southern Kurdish and the dialect of Mūkri in Central Kurdish, izāfa is dropped in certain structures which can of interest for further investigation. For example, in a definite phrase marked with *aka*, Mūkri speakers tend to put an izāfa only on the adjacent adjective to the head noun while the non-adjacent ones do not host an izāfa.

# Bibliography


Aronoff, M. (1994). *Morphology by Itself: Stems and Inflectional Classes*. Linguistic inquiry monographs. Niemeyer.

Aronoff, M. (2016). Unnatural kinds. In Luís, A. and Bermúdez-Otero, R., editors, *The Morphome Debate*, chapter 2, pages 11–32. Oxford University Press.

Barker, C. (2004). Possessive weak definites. *University of Massachusetts Occasional Papers*.

Bedir Khan, E. D. and Lescot, R. (1986). *Kurdische Grammatik. Kurmancî Dialekt*. Verlag für Kultur und Wissenschaft. (Translation of the French original from 1970).

Bergsma, S., McNamee, P., Baghdouri, M., Fink, C., and Wilson, T. (2012). Language identification for creating language-specific twitter collections. In *Proceedings of the Second Workshop on Language in Social Media*, pages 65–74. Association for Computational Linguistics.

Bermúdez-Otero, R. and Luís, A. (2016). A view of the morphome debate. In *The Morphome Debate*, chapter 12, pages 309–340. Oxford University Press.

Boas, H. and Sag, I. (2012). *Sign-based Construction Grammar*. CSLI Lecture Notes Lecture Notes. CSLI Publications/Center for the Study of Language and Information.

Booij, G., Lehmann, C., Mugdan, J., Kesselheim, W., and Skopeteas, S. (2000). *Morphologie / Morphology. 1. Halbband*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK). De Gruyter.

Cinque, G. (2010). The syntax of adjectives. a comparative study. *MITPress, Cambridge.*, LI Monograph(57).

den Dikken, M. (2006). *Relators and Linkers The Syntax of Predication, Predicate Inversion, and Copulas*. MIT Press.

Ghomeshi, G. (1997). Non-projecting nouns and the ezafe construction in persian. natural language and non-projecting nouns and the ezafe construction in persian. *Natural Language and Linguistic Theory*, 15(4):729–788.

Glosbe (2018). Glosbe dictionary.

Haider, H. and Zwanziger, R. (1984). Relatively attributive. the 'ezāfe'-construction from old iranian to modern persian. In of English, I., Linguistics, I., Fisiak, J., Instytut filologii angielskiej (Poznań, P., Syntax, I., im. Adama Mickiewicza w Poznaniu, U., and English, U., editors, *Historical Syntax*, Studies and monographs. Mouton.

Haig, G. (2011). Linker, relativizer, nominalizer, tense-particle. In Foong Ha Yap, Karen Grunow-Hårsta, J. W., editor, *Nominalization in Asian Languages: Diachronic and typological perspectives*, volume 96 of Typological Studies in Language, pages 363–390. John Benjamins Publishing.

Haig, G. and Matras, Y. (2002). Kuurdish linguistics: a brief overview. *STUF-Language Typology and Universals*, 1(55):3–14.

Hincha, G. (1961). Beitrage zu einer morphemlehre des neup. *Der Islam*, (37):136–201.

Kahnemuyipour, A. (2003). Syntactic categories and persian stress. *Natural Language and Linguistic Theory*, 21(2):333–379.

Kahnemuyipour, A. (2014). Revisiting the persian ezafe construction: a roll-up movement analysis. *Lingua*, 150:1–24.

KAL (2018). Kurdish language.

Karimi, S. and Brame, M. (2012). A generalization concerning the ezafe construction in persian. *Linguistics Analysis*, 38:11–143.

Karimi, Y. (2007). Kurdish ezafe constructions: Implications for dp structure. *Lingua*, 117(12):2159–2177.

Klavans, J. (1985). The independence of syntax and phonology in cliticization. *Language*, 61(1):95–12.

Larson, R. (1991). The projection of dp (and degp). *Unpublished manuscript, Stony Brook University*.

Larson, R. and Yamakido, H. (2005). Ezafe and the deep position of nominal modifiers. *Presented at the Bacelona Workshop on Adjectives and Adverbs*.

Lass, R. (1990). How to do things with junk: exaptation in language evolution. *Journal of Linguistics*, 26(1):79–102.

Lazard, G. (1966). L'enclitique nominal -i en persan : un ou deux morphèmes? *Bulletin de la Société de Linguistique de Paris*, (61):249–264.

Löbner, S. (1985). Definites. *Journal of Semantics*, 4(4):279–326.

Löbner, S. (2011). Concept types and determination. *Journal of Semantics*, 28(3):279–333.

Mackenzie, D. N. (1961). *Kurdish Dilect Studies*. Oxford University Press.

Maiden, M. (1992). Irregularity as a determinant of morphological change. *Journal of Linguistics*, 28(2):285–312.

Malmasi, S. (2016). Subdialectal differences in sorani kurdish. *In Proceedings of the VarDial Workshop. Osaka, Japan.*

McCaurus, E. (2012). Kurdish. In Windfuhr, G., editor, *The Iranian Languages*, chapter 10. Routledge Series, 2nd edition.

Michaelis, L. (2009). Sign-based construction grammar. In Heine, B. and Narrog, H., editors, *The Oxford Handbook of Linguistic Analysis*, chapter 7. Oxford University Press.

Michaelis, L. A. (2006). Cosntruction grammar. In Brown, K., editor, *Encyclopedia of Language and Linguistics*, volume 3, pages 73–84. Elsevier Science.

Nerwiy, H. K. T. (2012). *The Republic of Kurdistan, 1946*. PhD thesis, Leiden University.

Poesio, M. (1994). Weak definites. In *Semantics and Linguistic Theory*, volume 4.

Pollard, C. and Sag, I. (1994). *Head-Driven Phrase Structure Grammar*. Head-driven Phrase Structure Grammar. University of Chicago Press.

Rebuschi, G. (2002). Coordination et subordination. deuxième partie : vers la co-jonction généralisée. *la Société de Linguistique de Paris*, XCVII:37–94.

Sag, I. (2012). Sign-based construction grammar: An informal synopsis. In Boas, H. and Sag, I., editors, *Sign-based Construction Grammar*, CSLI Lecture Notes, chapter 3. CSLI Publications/Center for the Study of Language and Information.

Sag, I., Wasow, T., and Bender, E. (2003). *Syntactic Theory: A Formal Introduction*. Stanford: CSLI Publications, 2 edition.

Salih, R. R. (2014). *A comparative study of English and Kurdish connectives in newspaper opinion articles*. PhD thesis, University of Leicester.

Samiian, V. (1983). *Origins of phrasal categories in Persian: An X-bar analysis*. PhD thesis, Ph.D. dissertation, University of California at Los Angeles.

Samiian, V. (1994). The ezafe construction: Some implications for the theory of x-bar syntax. *Persian Studies in North America. Bethesda, MD: Iranbooks*, pages 17–41.

Samvelian, P. (2007a). A lexical account of sorani kurdish prepositions. *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar Stanford Department of Linguistics and CSLI's LinGO Lab*, pages 235–249.

Samvelian, P. (2007b). A (phrasal) affix analysis of the persian ezafe. *Journal of Lingusitics*, 43(3):605–645.

Samvelian, P. (2007c). What sorani kurdish absolute prepositions tell us about cliticization. *Texas Linguistic Society IX: The Morphosyntax of Underrepresented Languages*, page 265.

Samvelian, P. (2008). *Aspects of Iranian Linguistics*, chapter The Ezafe as a head-marking inflectional affix: evidence from Persian and Kurmanji Kurdish, pages 339–361. Cambridge Scholars Publishing, Newcastle upon Tyne.

Sheykh Esmaili, K. and Salavati, S. (2013). Sorani kurdish versus kurmanji kurdish: An empirical comparison. *ACL*, pages 300–305.

Taghipour, S. (2017). Laki verbal inflection. Master's thesis, University of Kentucky.

Thackston, W. (2006). *Sorani Kurdish: A reference Grammar with Selected Readings*. Renas Media.

TKD (2018). The kurdish dictionary.

Zahedi, K. and Mehrazmay, R. (2011). Definiteness in sornai kurdish and english. *Dialectologia: revista electronica*, (7):129–157.

Zribi-Hertz, A. and Jean-Louis, L. (2013). From noun to name: definiteness marking in modern martinikè. In Zribi-Hertz, P. C. H. . A., editor, *Crosslinguistic studies on Noun Phrase structure and reference*, volume 39 of *Syntax and Semantics*, pages 269–315. Brill, https://halshs.archives-ouvertes.fr/halshs-01060056.

# Vita

Ali Salehi

## Education

2016-2018 M.A. in Linguistic Theory and Typology, University of Kentucky, Lexington, Kentucky. (In progress)

2007-2011 B.A. in English Language and Literature, Razi University, Kermanshah, Iran.

## Professional Positions

2016-2018 Teaching Assistant, Primary Instructor, Department of Linguistics, University of Kentucky, Lexington, Kentucky.