Bamberg Studies in Kurdish Linguistics 1

Songül Gündoğdu, Ergin Öpengin, Geoffrey Haig,
Erik Anonby (eds.)

Current issues in Kurdish linguistics

University
of Bamberg
Press

1 Bamberg Studies in Kurdish Linguistics

Bamberg Studies in Kurdish Linguistics

Volume 1

University
of Bamberg
Press
2019

Current issues in Kurdish linguistics

Songül Gündoğdu, Ergin Öpengin, Geoffrey Haig,
Erik Anonby (eds.)

University
of Bamberg
Press
2019

Bibliographische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Informationen sind im Internet über http://dnb.d-nb.de/ abrufbar.

Diese Veröffentlichung wurde im Rahmen des Elite-Maststudiengangs „Kulturwissenschaften des Vorderen Orients“ durch das Elitenetzwerk Bayern gefördert, einer Initiative des Bayerischen Staatsministeriums für Wissenschaft und Kunst.

Herstellung und Druck: Digital Print Group, Nürnberg
Umschlaggestaltung: University of Bamberg Press

ISSN: 2698-6612
ISBN: 978-3-86309-686-1 (Druckausgabe)
eISBN: 978-3-86309-687-8 (Online-Ausgabe)
URN: urn:nbn:de:bvb:473-opus4-558751
DOI: http://dx.doi.org/10.20378/irbo-55875

Acknowledgements

This volume contains a selection of contributions originally presented at the *Third International Conference on Kurdish Linguistics* (ICKL3), University of Amsterdam, in August 2016. We are very grateful to the organizers of the conference for providing the impetus for this volume, to a number of anonymous referees for their expert input in the reviewing procedure, and of course to all our authors who have patiently cooperated throughout the lengthy editorial process. In particular, we would like to thank Jenny Herzky for supervising the technical issues involved in formatting a complex manuscript, and Hannah Sarrazin and Maria Vollmer for assistance in layout and proof-reading. We would also like to express our gratitude to the Department of General Linguistics in Bamberg for financial support, and to the Elitenetzwerk Bayern, who finance the Bamberg MA-programme "Cultural Studies of the Middle East" and who made a generous contribution to the production costs of this volume.

The Editors

Contents

Foreword to the new series *Bamberg Studies in Kurdish Linguistics* (BSKL)

Despite the chronic shortage of institutional support, research in Kurdish linguistics continues to thrive. Among recent developments, one can mention the *International Conference on Kurdish Linguistics*, which began as an informal workshop in Bamberg in 2013 and has since grown into a regular international conference series, or the *Database of Kurdish Dialects*, the first large-scale web-accessible dialect survey of Kurdish, hosted at the University of Manchester. What remains a desideratum, however, is a platform for disseminating quality research in a readily accessible format.

The new series *Bamberg Studies in Kurdish Linguistics* (BSKL) aims at filling this gap by providing a publication forum for high-quality research on Kurdish linguistics that is open access, hosted at an established research institution, and committed to high standards of scientific excellence. BSKL adopts a broad stance on what can be considered "Kurdish linguistics", in terms of the languages in focus, the range of topics, and the format of the volumes. The primary criterion is scientific excellence, and the editors are committed to ensuring high standards. It seems very appropriate that the first volume in the series is a selection of contributions to the *International Conference on Kurdish Linguistics*, held in Amsterdam in 2016, which provides a vivid cross-section of contemporary research in Kurdish linguistics.

It has taken several years of planning before we could finally inaugurate the series, and we are extremely grateful to a number of people for their patience and support during that time, in particular the staff at Bamberg University Press, who agreed to the volume conception several years ago and have continued to support it since. Finally, I would like to thank the co-editors of the series, Erik Anonby, Ergin Öpengin and Ludwig Paul, for collaborating with me on the new series.

Bamberg, 5th July 2019

Geoffrey Haig, editor-in-chief of BSKL

1
Introduction

Songül Gündoğdu, Ergin Öpengin, Geoffrey Haig & Erik Anonby

The history of Kurdish linguistics can be traced back to the 17th and 18th centuries with the works of Kurdish authors, who worked within the widespread Arabic-centred tradition of grammar writing (cf. Leezenberg 2014). This was followed by works in the Orientalist tradition, the first one being Maurizio Garzoni's notable grammar and vocabulary of Kurmanji Kurdish published in 1787. In the subsequent two centuries, up until the mid-20th century, a large number of works describing the grammar of particular dialects appeared, and often vocabulary samples and construction types from Kurdish were also incorporated into comparative studies by scholars of Indo-European grammar. Most of those grammars are modeled after the categories distinguished in Latin or those of the working language of the orientalist author. Noteworthy are also the two "sensibly prescriptive" grammars of Sorani and Kurmanji Kurdish in the second quarter of 20th century, the first one by Tawfiq Wahbi (1929) and the second one by Djeladet Ali Bedir-Khan (Celadet Alî Bedirxan) with Roger Lescot (1970), itself based on a series of articles in the journal *Hawar* that were published between 1932 and 1943.

Work in Kurdish linguistics – in its modern sense – began only in the 1950s, with the studies of such authors as C. J. Edmonds (e.g. 1955), David N. MacKenzie (e.g. 1961), and Ernest McCarus (e.g. 1958) providing principled descriptions of Kurdish grammar as well as accounts of particular phenomena in Kurdish morphology and syntax. However, with extremely limited institutional support the research on Kurdish remained sporadic and constrained. It was only with the new millennium, when the visibility of the Kurds on the international scene increased and sovereign states governing Kurdish societies eased restrictions on the use of and work on Kurdish, that there has been increased interest in Kurdish linguistics in international academia. Over the last two decades, many young scholars, including those

at universities in Kurdish-speaking regions of Iraq, Turkey and Iran, have had the chance to pursue research projects for their post-graduate studies on Kurdish.

However, collective venues such as conferences and edited volumes or journals dedicated to Kurdish linguistics have been scarce. The most important event along these lines was the *First Kurdish Linguistics Conference* organized at the University of Kiel, Germany, in 2000, the contributions of which were collected in a special issue of the journal *Language Typology and Universals* (2002). The conference was not continued as a recurrent event, but research in Kurdish linguistics began to figure more frequently in other topically broader conferences, particularly at the biennial *International Conference on Iranian Linguistics*. However, a workshop focusing specifically on "Variation and Change in Kurdish" was held in 2013 at the University of Bamberg and a selection of its contributions were edited in a special issue of the journal *Kurdish Studies* (2014). The introductory article there (Haig & Öpengin 2014) summarizes the current topics, trends, and desiderata in the field. The workshop was continued in the form of the *International Conference on Kurdish Linguistics* (ICKL), which gathered first at Mardin Artuklu University (2014) and subsequently at the University of Amsterdam (2016) and the University of Rouen (2019). In a short span of time this conference has turned into a lively venue for the presentation of new research, exchange of ideas and establishment of collaborations among scholars from different generations and countries. The present volume, *Current Issues in Kurdish Linguistics*, brings together a robust selection of contributions presented at the Amsterdam conference. However, it is more an edited volume than a typical "proceedings", since the presentations were solicited in the format of research articles, and a rigorous peer-review and editorial procedure was implemented – hence the extended time frame in its appearance as a book.

The volume contains ten contributions which span the field of Kurdish linguistics in a geographic as well as a topical sense. Along with several works on Kurmanji (Northern Kurdish) and Sorani (Central Kurdish), two chapters shine light on the lesser-known Southern Kurdish language area. Other studies are comparative and treat the Kurdish language area as a whole. The linguistic approaches of the authors are a mixture of formal and typological perspectives, and cover topics ranging from geographical distribution and variation to phonology, morphosyntax, discourse structure, and sociolinguistics.

In the first chapter, Erik Anonby, Masoud Mohammadirad and Jaffer Sheyholislami provide a first detailed and comprehensive picture of language

distribution in Kordestan Province, which is one of four provinces in Iran where Kurdish is the main spoken language. Their research has been carried out in the context of the *Atlas of the Languages of Iran* (ALI) research programme (see Anonby & Taheri-Ardali, et al. 2015–2019). The article presents the methodology and results of new data collection, consisting of local place names and language distribution data, combined with existing data sets and mapped out to the level of each settlement. The authors show that the language situation in Kordestan Province is diverse, with six important high-level varieties represented: Central Kurdish, Southern Kurdish, Hawrami, Turkic, Persian and Aramaic. Moreover, most of these varieties display substantial internal variation, and are thus classified into major sub-varieties. Among the high-level varieties, the authors specifically focus on the labelling and internal classification of Central Kurdish and Southern Kurdish for which the dialect situation presents a number of complexities.

The existence of pharyngeal and emphatic articulations in Kurdish has long been considered one important feature of Kurdish, setting it apart from most other Iranian languages on one hand and, on the other hand, pointing to intense contact with Semitic languages, especially Arabic. However, a theoretically-informed treatment of such observations, focusing on phenomena specific to Kurdish, has been missing. Daniel Barry, in his chapter on pharyngeal sounds in Kurdish, provides a detailed account of these sounds both in Arabic loans and the native component of the lexicon, also defining and characterizing the phonological environments where they typically occur. He argues that while pharyngeal phonemes have been introduced into Kurdish through Arabic loans, their propagation into the native lexicon in Kurmanji is the result of an internal phonological process that is modulated by speakers' familiarity with the phonetics of Arabic pharyngeals. The phonological process in question builds upon the association of pharyngealized vowel phonemes with a subset of consonants, particularly labials, and constraints determined by the phonotactics of the language. Barry's analysis is revealing in accounting for the particular phonological environments in which pharyngealization in the native component occurs, and in accounting for different historical layers in the treatment of pharyngeal articulations in Kurdish.

In her chapter on the dialectology of Southern Kurdish, Sara Belelli sheds light on major methodological questions in approaching Southern Kurdish, and addresses core issues of terminology. The article meticulously reviews Fattah's (2000: 9) proposed dialect classification alongside other

existing sources, and complements it with the author's own documentation activities in Kermanshah Province and elsewhere in the Southern Kurdish-speaking area. The author delineates the geographic extension of the Southern Kurdish continuum from Kordestan Province, across Kermanshah and south towards the Laki-speaking regions of Lorestan and east Ilam. Laki-Kermānshāhi and Kordali function as transitional links to Laki and Lori respectively. Belelli concludes, however, that more in-depth study is required before any assertion on genetic affiliations or the direction of contact-induced variation in border areas can be made.

"Asymmetries in Kurmanji morphosyntax", by Songül Gündoğdu, scrutinizes the status of postverbal goals and certain adpositional phrases in Kurmanji Kurdish in order to understand why some phrases appear in the immediate postverbal position and why certain object-like constituents are adpositional, unlike direct objects in this language. The discussion in this article reveals that the morphological realization of the constituents (case vs. adposition) and their linear ordering (preverbal vs. postverbal) in a Kurmanji clause are sensitive to the correlation between verb meaning and event type: structural participants are realized with case morphology while constant participants are introduced with adpositions. Furthermore, the author argues that the reason Kurmanji makes a distinction in the linear ordering of structural participants is indeed a word-order property (Verb-Goal) retained from proto-Kurdish and further constrained by the morphosyntactic properties of this language.

Geoffrey Haig's contribution, "Debonding of inflectional morphology in Kurdish and beyond", defines and discusses examples of "debonding" in Kurdish and other West Iranian languages, including Balochi and Tatic varieties. Debonding involves a loosening of the distributional and phonological attachment between an inherited inflectional affix and its base. The consequences of debonding may be the intrusion of additional morphology between the affix and base, or the analogical extension of the debonded affix to hosts of other classes that were not previously associated with that morpheme (e.g. from nouns to pronouns). Debonding runs counter to the expectations of grammaticalization theory, according to which inflectional morphology is predicted to either remain as such, or erode to zero. There are obvious connections to the shift from affix to clitic observed for person indexing morphology in Öpengin's chapter on Central Kurdish in this volume. Haig's chapter also considers the unexpected sequences of definiteness and number morphemes in Southern and Central Kurdish, but suggests that, despite certain superficial similarities, this is not an example of

debonding, but arises through the unusual source of definiteness marking from diminutives.

Geoffrey Haig and Baydaa Mustafa investigate patterns of language use and language attitudes among Bahdini Kurdish speakers in the multilingual city of Duhok in Iraqi Kurdistan. The Bahdini variety of Northern Kurdish is sociolinguistically doubly disadvantaged: within the context of Iraqi Kurdistan, it is the smaller and less prestigious variety in comparison to Sorani, and in the context of Northern Kurdish it is relatively peripheral, isolated from the pan-national Botan standard by certain dialectal features, and the use of the Arabic script. The authors present the results of structured interviews and lexical retention tasks with more than 100 speakers, from three generations. The results reveal both age- and gender-related differences, indicating that awareness of Kurdish language issues is more strongly reflected in the speech of younger generations, though this does not necessarily match reported attitudes towards Kurdish as the language of education. In particular, the age cohort of 31- to 50-year-old adults exhibits variable attitudes, with a strong effect of gender, possibly reflecting this generation's first-hand experience in the difficult formative phases of Kurdish language education. This is the first empirical study of this scale from the Duhok region, and provides a basis and impetus for further research on one of the most complex multi-lingual settings in contemporary Kurdish society.

In her corpus-based research on the development of conjunctions in Kurmanji Kurdish, Annette Herkenrath takes a synchronic look at the transitional area between clausal and NP-level patterns of junction, based on a corpus of academic writings published in Kurmanji. The article focuses on lexical nouns with a temporal meaning (i.e. temporal nouns – TN) such as *dem* 'time, period', *gav* 'moment, time, step', *wext* 'time, period, season' and *çax* 'time, age, period, era'. Since these temporal words function as nouns, junctors, and adpositions depending on their syntactic environment, they can flexibly change roles between lexical noun and subordinating junctor. The author demonstrates that finite subordinate clauses appear at one end of a scale of TN modifiers, after nouns, action nouns, verbal nouns and participles, whereas clause-embedding TNs may express up to two functional categories associated with the NP area. Moreover, at the intersection of these two continua, constructions can be observed to subtly transit into and out of "nouniness" at both levels simultaneously.

"Kurdish *-râ* as an Anti-Actor marker", by Gholamhossein Karimi-Doostan and Fatemeh Daneshpazhouh, investigates the semantic and syntactic roles of the *-râ* morpheme in Sorani Kurdish. The authors adopt the framework of

Role and Reference Grammar, specifically the notion of Actor and Undergoer macroroles. Discussing the semantic and syntactic functions of *-râ* in various constructions, the article demonstrates that this morpheme appears in non-active clauses which lack Actor external arguments (i.e. the DPs with Actor roles) at the syntax level. In other words, it shows up when an Actor role is semantically present, but syntactically absent. Therefore, the authors argue that *-râ* is an Anti-Actor morpheme, as its presence leads to the absence of arguments with Actor roles.

One of the major undertakings in the field of Kurdish linguistics has been the collaborative project "Structural and typological variation in the dialects of Kurdish", led by Yaron Matras between 2011–2017, and based at the University of Manchester.[1] Preliminary findings of this project pertaining to the distribution of structural features and dialect geography and classification are presented in a chapter by Yaron Matras. Matras presents the background to the project, the methodology of data collection, data types and the processing of the data. His account builds upon previous work on Kurdish dialect classifications and is based on data from over 150 locations in Kurdish-speaking regions of Turkey, Iran, Iraq, and Syria. Appealing to the principle that it is innovation that creates differences among related varieties, Matras sets out to reconstruct the layers of structural innovation and investigate the extent of their diffusion in geographical space. Since the extent to which individual innovations spread are variable, there are no pre-determined dialect boundaries in the survey, which is also a methodological novelty of the project in the study of Kurdish dialects. Matras summarizes project findings on the well-known "great divide" between Northern Kurdish and Central Kurdish before presenting in detail the innovations that characterize sub-areas or epicentres of the northern and southern dialect clusters of respective varieties. He further elaborates on retention zones, especially the archaic convergence zone between Northern and Central Kurdish to the north of Erbil province. A welcome addition to the dialectology of Kurdish in Matras' paper is the three-way classification of Kurdish dialects in Syria. Matras concludes with a detailed hypothesis on the relationship of the Northern and Central Kurdish dialect clusters and the stages through which the innovation and retention zones have been shaped, stemming from the current complexity of intersecting isoglosses across a relatively large space.

[1] For further information on the project, see:
`http://kurdish.humanities.manchester.ac.uk/` (accessed 5 July 2019).

Finally, in his contribution on the interactions of two categories of person markers in Central Kurdish, Ergin Öpengin considers several theoretical and language-internal problems around combinations of pronominal clitics and verbal agreement suffixes. In Central Kurdish, in certain morphosyntactic constellations elements from these two person marker categories occur in sequence (each one indexing a separate grammatical relation). In such combinations, the pronominal clitic occurs before the verbal agreement suffix, except with a third person singular pronominal clitic which regularly comes after the verbal agreement suffix. This poses a double problem: (i) being syntactic entities, pronominal clitics are expected to occur external to morphological verbal affixes; and (ii) the idiosyncratic behaviour of third person singular pronominal clitic violates the paradigmatic structure of the person marker categories. In his analysis nested in Prosodic Phonology, Öpengin looks into the facts of lexical stress in Central Kurdish, and shows that the verbal agreement person markers – which are historically "affixes" – do not phonologically integrate into the verb stem to which they attach. That is, they do not form a phonological/prosodic word with their host, and as such are clitic-like in this regard. The observed order in combination, with pronominal clitics preceding the verbal agreement person markers, then, follows the general second-position placement principle of pronominal clitics in Central Kurdish. As for the idiosyncratic ordering with a third person singular pronominal clitic, this is seen as an exceptional case of "identity avoidance" that results from the constraints that require the forms in sequence to preserve their distinct morpho-phonological identity and effectively express the encoded morpho-syntactic function (e.g. grammatical relations, argument roles).

The present edited volume is the first of its kind in bringing together contributions from a relatively large number of linguists, working in a diverse range of frameworks and on different aspects and varieties of Kurdish. As such, it attests to the increasing breadth and sophistication now evident in Kurdish linguistics, and is a worthy launch for the new series *Bamberg Studies in Kurdish Linguistics* (BSKL).

References

Anonby, Erik & Taheri-Ardali, Mortaza, et al. (eds.). 2015–2019. *Atlas of the Languages of Iran*. Ottawa: Geomatics & Cartographic Research Centre, Carleton University. (`http://iranatlas.net/`) (Accessed 2019-06-15).

Bedir-Khan, Emir Djeladet & Lescot, Roger. 1970. *Grammaire kurde (dialecte kurmandji)*. Paris: Maisonneuve.

Edmonds, Cecil J. 1955. Prepositions and personal affixes in Southern Kurdish. *Bulletin of the School of Oriental and African Studies* 17(3). 490–502.

Fattah, Ismail Kamāndār. 2000. *Les dialectes kurdes méridionaux: Étude linguistique et dialectologique* (Acta Iranica 37). Leuven: Peeters.

Garzoni, Maurizio. 1787. *Grammatica e vocabulario della lingua kurda*. Roma: Nella Stamperia della Sacra Congregazione di Propaganda Fide.

Haig, Geoffrey & Öpengin, Ergin. 2014. Kurdish: A critical research overview. *Kurdish studies* 2(2) [Special issue: *Kurdish linguistics: Variation and change*]. 99–122.

Leezenberg, Michael. 2014. Elî Teremaxî and the vernacularization of Medrese learning in Kurdistan. *Iranian studies* 47(5). 713–733.

MacKenzie, David Neil. 1961. *Kurdish dialect studies*. Vol. 1. London: Oxford University Press.

McCarus, Ernest N. 1958. *A Kurdish grammar: Descriptive analysis of Kurdish of Suleimaniya, Iraq*. New York: American Council of Learned Societies.

Wahbi, Tawfiq. 1929. *Destûrî zimanî kurdî*. Baghdad: al-Hadith.

2
Kordestan Province in the *Atlas of the Languages of Iran*: Research process, language distribution, and language classification

Erik Anonby, Masoud Mohammadirad & Jaffer Sheyholislami

Abstract: Kordestan is one of four provinces in Iran where Kurdish is the main spoken language. A small number of studies of specific language varieties in Kordestan Province have appeared, and the province is featured as part of several general regional or country-wide maps of language distribution. Until now, however, no systematic study on the language situation in this province has been published. The present paper, which seeks to address this gap, provides an account of the research currently being carried out on Kordestan Province of Iran in the context of the *Atlas of the Languages of Iran* (ALI) research programme. After introducing the Atlas project and the research team for Kordestan, we look at the role of existing data sources in the Atlas, including the production of a background map and an online bibliography of language-related resources. The main portion of the paper deals with the collection of new data, consisting of local place names and language distribution data, combined with existing data sets and mapped out to the level of each settlement. The results of our study show that the language situation in Kordestan Province is more diverse than often assumed, with six important high-level varieties represented: Central Kurdish, Southern Kurdish, Hawrami, Turkic, Persian and Aramaic. Most of these varieties also show significant internal variation, as shown by our inventory and initial classification of all major subvarieties. The study concludes with reflections on the importance of a fine-grained and systematic approach to investigating the language situation, the limitations of this type of large-scale study, and possibilities for further research that refines and builds on the findings presented here.

1 Introduction[1]

Kordestan Province of Iran, with a population of just over 1.6 million (ISC 2016), is one of four provinces in Iran – along with Kermanshah, Ilam and West Azerbaijan[2] – where Kurdish is the main spoken language. Located in the north-west part of the country, Kordestan is bounded to the north by Central Kurdish-speaking regions of West Azerbaijan Province of Iran; the primarily Turkic-speaking provinces of Zanjan and Hamadan to the east; the Southern Kurdish-dominated Kermanshah Province to the south; and Central Kurdish-speaking areas in the Kurdistan Regional Government of Iraq to the west.

A handful of studies covering specific language varieties in Kordestan Province have been published (Fattah 2000; Sohrābi & Serish Ābādi 2009; Kordzafarānlu Kāmbuziā & Sajjadi 2013; Sajjadi & Kordzafarānlu Kāmbuziā 2014), along with a larger number of MA theses (e.g. Rezāi 1996; Teymuri 1998; Hasanzādeh 1999; Khaliqi 2001; Mohammadi 2002). The province is also featured as part of several overview maps of Kurdish (Hassanpour 1992, revised in Haig & Öpengin 2014 and Sheyholislami 2015; Izady 1998; Matras & Koontz-Garboden 2017, and related studies including Anonby (forthcoming)) and general maps of language distribution across Iran (*Atlas Narodov Mira* 1964, *TAVO* 1988, *Compendium* 1989, Izady 2006–13, *Irancarto* 2012). It is commonly assumed that Central Kurdish is the characteristic language of the province as a whole (e.g., "Kurdistan Province" in *Wikipedia*[3]). However, no detailed or systematic study has been published which focuses on the language situation in this province, although Khādemi's (2002) MA thesis is an important

[1] An earlier version of this paper was presented by the authors at the *3rd International Conference on Kurdish Linguistics* (ICKL3), University of Amsterdam, August 25–26, 2016. The authors wish to acknowledge the contributions of SSHRC (Social Sciences and Humanities Research Council of Canada), Carleton University, Universität Bamberg, the Alexander von Humboldt Foundation, and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 665850. We are grateful for insightful and detailed feedback from an anonymous referee and the volume editors.

[2] As is the case for most other parts of Iran, there are no reliable or detailed data on language distribution in West Azerbaijan Province. A number of districts in the province are majority Azerbaijani-speaking, including the capital city of Orumieh (Urmia). Because of this – and perhaps also because of the province's name – it is often assumed that Azerbaijani is the main language of the province as a whole. However, our own preliminary investigations of this topic, which are based on district-by-district calculations, together with the maps found in *Irancarto* (2012), suggest that Kurdish may in fact be the mother tongue of a slight majority of the province's population.

[3] Available at `https://en.wikipedia.org/wiki/Kurdistan_Province` (accessed April 23, 2019).

effort in this direction. The present study seeks to address this gap in the literature.

This paper provides an account of the research currently being carried out on Kordestan Province of Iran in the context of the *Atlas of the Languages of Iran* (ALI) research programme (Anonby & Taheri-Ardali, et al. 2015–2019; Anonby et al. 2019). It consists of an outline of the research process and the findings that have been generated so far. Our investigation, which has been conducted to the level of each settlement, provides a first detailed and comprehensive picture of language distribution in Kordestan Province.

After introducing the Atlas research team for Kordestan Province, we look at the incorporation of existing data sources, including key demographic and geographic sources, the production of a background map for mapping language in Kordestan Province, and a continuously developed bibliography of linguistic and sociolinguistic references.

The main portion of the paper deals with the collection of new data, consisting of local place names and language distribution data. An extensive discussion of methodology precedes an inventory and analysis of the results for both of these topics. In the section on local place names, we show that this seemingly peripheral step in the research makes several important contributions to the overall objective of understanding the language situation. The core of the research treats language distribution and classification for the major varieties and subvarieties of Kordestan Province, investigated and mapped out to the level of each settlement. The results of our study show that the language situation is more diverse than often assumed, with six important high-level varieties represented: Central Kurdish, Southern Kurdish, Hawrami, Turkic, Persian and Aramaic. Special attention is given to the labelling and internal classification of Central Kurdish and Southern Kurdish, for which the dialect situation presents a number of complexities.

The study concludes with reflections on the importance of a fine-grained and systematic approach to investigating the language situation, limitations of this type of large-scale investigation, and possibilities for further research that refines and builds on the results of this research.

2 The *Atlas of the Languages of Iran* (ALI)

There have been a number of important efforts to map of the languages of Iran, but until now no language atlas, or even a comprehensive and detailed country-level language map, has been produced. As detailed in Anonby (2015),

this can be attributed to a variety of factors including the complexity of language situation; issues of logistics and project design; contrasting perspectives on language identity and distribution; limited dissemination of project results; and limited cooperation among scholars working toward this common goal.

After five years of planning, work on the *Atlas of the Languages of Iran* (ALI) (test version: `http://iranatlas.net`) began in earnest when seed funding was obtained in 2014. In this research programme (Anonby et al. 2019), an atlas of the country's languages is being developed by an international team of over 80 volunteer scholars and students. This atlas, which includes each of Iran's some 60,000 cities and villages, brings together existing publications and new data. It is capable of remote contributions by scholars and popular users and moderation of input by atlas editors. Because ALI brings together the work of many different people, it provides references to each data source, whether published work, collaborator field notes or user contributions. Fundamental to the purpose of the Atlas, it is designed to facilitate comparison of language distribution maps with maps based on attested linguistic forms (Anonby et al. 2016; Anonby & Sabethemmatabadi (In press)). ALI is being built by GCRC (Geomatics and Cartographic Research Centre): `https://gcrc.carleton.ca`) using the open-source Nunaliit Atlas Framework (`http://nunaliit.org`; GCRC 2013), which comes with a ready-made atlas template. The language mapping functionality developed in the present research programme, presented in Anonby et al. (In press), is continuously incorporated into the Nunaliit platform and made freely available to other scholars on GitHub (`https://github.com/GCRC/nunaliit`).

The present research on Kordestan Province has been carried out in the context of this larger research programme. It is the third major geographical focus in the Atlas, preceded by initial work on the provinces of Hormozgan (Mohebbi Bahmani et al. 2015) and Chahar Mahal va Bakhtiari (Taheri-Ardali et al. 2015; Anonby & Taheri-Ardali 2018; Taheri-Ardali & Anonby 2019); and it is being followed by detailed studies of language in Bushehr (research team: Nemati et al.), Kermanshah (Fattahi et al.) and Ilam (Aliakbari et al. 2014, Gheitasi et al.). With each new province, we are streamlining the research process and methodology for collection, analysis and presentation of the data.

3 Research process for Kordestan province in ALI

In this section of the paper, we provide an overview of key aspects of the research process. First, we introduce the members of the research team who have contributed to this study. We then look at the use of existing data sources to provide a foundation and context for research on languages distribution in Kordestan Province. Finally, we introduce the ways in which this study contributes to an understanding of the language situation through the collection of new data through field research.

3.1 The Atlas team for Kordestan Province

The present research on Kordestan Province has been carried out by a large and diverse research team within the context of the ALI research programme. Researchers who contributed to the current study, listed according to their specific roles in the research on Kordestan Province in the Atlas, are as follows:

Erik Anonby (Carleton/Bamberg/GCRC)
Project leader, Atlas editor, data consistency

Masoud Mohammadirad (Hamadan/Sorbonne Nouvelle Paris III)
Project consultant, field researcher, settlement localization

Jaffer Sheyholislami (Carleton)
Project consultant, bibliography, settlement localization

Mortaza Taheri-Ardali (IHCS/Shahrekord)
Atlas team coordinator

Fraser Taylor (Carleton/GCRC)
Project co-investigator

Amos Hayes (GCRC)
Geographic information technologist

J.-P. Fiset (GCRC)
Atlas programming

Robert Oikle (Carleton/GCRC)
Atlas design, map production

Adam Stone (Carleton)
Settlement localization

Sheema Rezaei (Carleton)
Settlement localization

Emily Wang (Carleton)
Settlement localization

Pegah Nikravan (Carleton)
Settlement localization

Parisa Sabethemmatabadi (Carleton)
Settlement localization

Ali Ghaharbeighi (Carleton)
Settlement localization

Partow Mohammadi (independent scholar)
Settlement localization

Ayat Tadjalli (Carleton)
Settlement localization

Nima Kiani (independent scholar)
Settlement localization

Laura Salisbury (Carleton/GCRC)
Map production

Ronak Moradi (Razi-Kermanshah)
Bibliography

3.2 Incorporation of existing data sources

The initial phase of the current research involved the collection and assembling of existing data sources. First, we developed a background map of Kurdistan Province designed specifically for language mapping. Using open-access data (SRTM 2014), Amos Hayes, Robert Oikle and Laura Salisbury constructed a chromatically neutral relief background, with administrative borders as a guide for map users (NCC 2014), onto which language distribution and linguistic data can be projected.

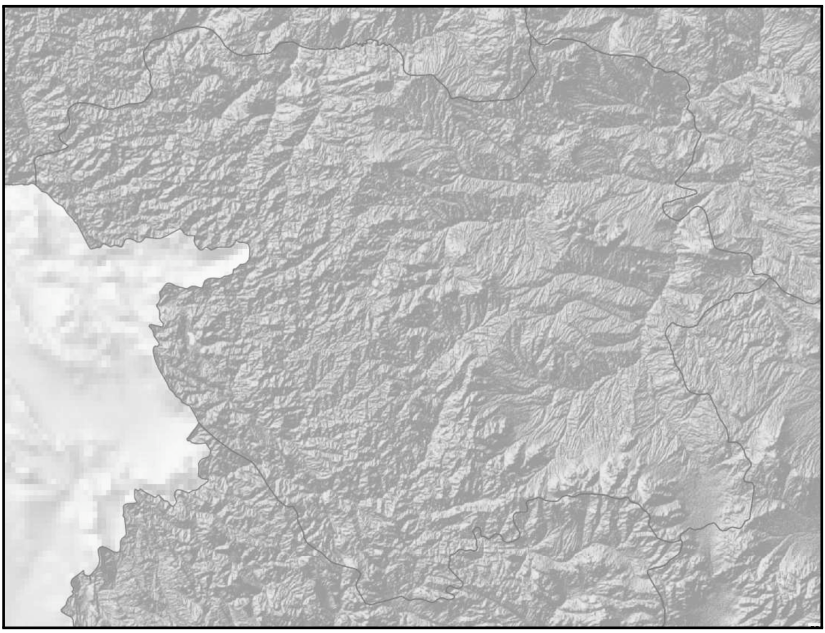

Figure 1: Background map for Kordestan Province
(Map design: Robert Oikle, GCRC. Source: `http://iranatlas.net`)

Secondly, we brought together settlement-related geographic data (NCC 2016, Roostanet 2016) and open-access demographic data for all populated places of Kordestan from the 2011 census of Iran (ISC 2011), which was the most recent available census data at the time of research. However, because the geographic data has not been made publicly available in tabular format, members of the research team spent several hundred hours reconstructing georeferenced (GPS) coordinates for each settlement.

The ALI bibliography brings together a third set of existing information sources for language in Kordestan Province. This annotated bibliography, which includes all works that address language distribution or provide linguistic data from Kordestan, has been assembled and is under continuous expansion by Jaffer Sheyholislami and Ronak Moradi. This is a slow and challenging task because many sources, whether commercial publications or academic works such as theses, have little or no presence on the internet.

This is especially true for studies written in Kurdish and Persian; many of these items are only available from bookstores and universities within the language area. While this bibliography is important in providing a foundation for our own work on Kordestan, as well as comprehensive referencing of the data in the Atlas, it is also valuable in its own right as resource for scholars, since it is the most complete repository of language-related materials for Kordestan Province.

3.3 Collection of new data: Language distribution and local place names

Building on the existing sources incorporated into the Atlas, the current research has entailed the systematic collection of preliminary data for all of the some 1800 settlements (i.e. cities, towns and villages) of Kordestan Province. For each settlement, we considered two issues: language distribution, and local pronunciations of the place names. The basic research questions we asked in relation to language distribution were as follows:

1. *What languages, and what subvarieties of these languages, are spoken as a mother tongue in this settlement?*
2. *In the case that more than one variety is spoken in the settlement, what is the estimated proportion of mother tongue speakers of each variety?*

For the topic of local place names, we asked:

What is/are the local name(s) of this place, as pronounced locally?

Field research on language distribution and local place names was carried out over a 6-month period in 2015 by Masoud Mohammadirad, with additional time spent analyzing and verifying the data. Because of the logistical impossibility of visiting nearly two thousand settlements, research was carried out through a network of participants from across the province. The assembled data is based on sources of three types:

1. Local knowledge of the field researcher. Mohammadirad was born and grew up in the city of Qorveh, in the south-east corner of Kordestan Province. He also worked for five years in the neighbouring district of Deh Golān. This background has given Mohammadirad an in-depth understanding of patterns in the language situation there and elsewhere in the Province.

2. Teachers at local schools in each region. Teachers are well-placed to contribute local place names and to provide assessments of language distribution because their students come from many different villages, and because they are themselves highly mobile within the regions where they work. Most teachers across Kordestan Province come from within the province, and often work in areas near to their communities of origin, so they are already familiar with the languages they encounter as well as sociolinguistic tendencies for language use across the province. Conversely, the fact that teachers have been in most cases assigned to schools outside of their community of origin, along with their higher education, adds an element of wider perspective and scientific rigour that is beneficial to the research process.

3. Additional sources. Whenever Mohammadirad or the teachers did not have detailed knowledge of local place names or language distribution for a particular village or area, they contacted people from the area under investigation to verify their own hypotheses and to fill in gaps in their knowledge of the situation.

After Mohammadirad's fieldwork was completed, the authors verified the reliability of the results through a careful joint review of the data to identify and address points of variation in the results; comparison with maps and other studies outside of the present project (Hassanpour 1992; Haig & Öpengin 2014; Sheyholislami 2015, etc.) to assess how closely our results lined up with those of these other studies; and additional, direct contacts with speakers in numerous geographic locations to resolve areas of ambiguity in the data.

On a practical level, this component of the research process has positive consequences beyond the specific research questions which are being addressed. Importantly, researchers who carry out this initial phase of the Atlas work or Kordestan Province are well-prepared for subsequent work in language data collection; they have already become familiar with the regions they will be investigating and the languages spoken there, and they have established a network of potential hosts for fieldwork.

It is imperative to recognize that, for logistical reasons, most of the language distribution and local place name research has been carried out indirectly, as described in the preceding paragraphs. This study in no way purports to be a census, with a highly trained researcher or research team travelling to each settlement to collect language distribution and local place name

data from all individuals, or even a representative cross-section of individuals from each place. In this respect, it is subject to the limitations that come with any large, non-census data set.

However, we have observed that, simply by asking research questions for each settlement, a first systematic, detailed picture of the language situation has emerged for the province as a whole. Crucially, this research raises new questions about relationships and differences between language varieties, and provides direction for subsequent collection of the kind of linguistic data that will help to address these questions more thoroughly.

Importantly, the very fact of making the data publicly available in the Atlas means that our findings can be critiqued and refined. Through a rigorous data collection and editing process, the research team has made every effort to ensure the reliability of all the data that has been collected, and to provide a reference for each piece of data. Still, with so much data, there are certainly oversights and areas for improvement. Because of this, the Atlas has been designed, and is already capable to receive, moderate and reference feedback on each piece of data. In this way, Atlas users who are familiar with a specific local situation will be able to assist the research team in improving the accuracy of the data.

In the following sections, we present topically detailed discussion and results from our study of language distribution and local place names.

4 Local place names: Methodology, significance and patterns

The collection and transcription of local place names has constituted an integral part of the research on Kordestan Province. For this portion of the study, Mohammadirad prepared a reference list of settlements using the Persian place names in the 2011 census list (ISC 2011), and romanized them according to the transcription conventions developed and posted on the Atlas project website (`https://carleton.ca/iran/transcription/`). He then asked respondents with local knowledge of each settlement the following question (repeated from Section 3.3 above): *What is or are the local name(s) of this place, as pronounced locally?*

Mohammadirad, who is not only a speaker and writer of Kurdish but also a linguist with experience in phonetic and phonological analysis of the language, transcribed the answers to this question, for each place, using a phone-

mic transcription system available at the same web page.[4] Each individual transcription has been double-checked for consistency by the Atlas editor. The Persian name, its romanized form, and the phonemic transcription of the local name, along with a reference to the source for the transcription, all appear on a page for each settlement which is reachable by clicking on the settlement in the Kordestan Province language distribution overview map (see Figure 2 in Section 5.2 below) or directly through an Atlas search.

Even though the additional step of collecting and transcribing local place names may seem peripheral to the enterprise of language mapping, we found that it has several indirect but important benefits to the process and eventual impact of the research.

First of all, it compels the researcher to locate individuals who are actually familiar with local place names, and this is a very specific, localized kind of data. Since people (whether researchers or speakers of the languages under investigation) tend to generalize tendencies about language distribution to whole areas without considering each settlement, this additional step ensures that the researcher will be in contact with people who have this very local knowledge of each settlement, before pursuing questions about language distribution.

Secondly, the featuring of local place names in the Atlas has the potential to strengthen the connection between the Atlas and its users, since people from a given settlement might be gratified at the public recognition of one local element of language and culture.

Thirdly, and closely tied to this, by enabling the production of maps with local names, the Atlas counterbalances the majority-language perspective inherent in all official maps of the region. In the end, it is the people living in a given settlement who use its name most frequently, and this element of local heritage deserves to be represented.

Finally, work on local names allows the researcher to become familiar with a diverse and representative (albeit limited) set of linguistic structures that are characteristic of each region within the province. In Atlas work on other

[4]The phonemic transcription system for local place names uses a minimum of complex phonetic symbols, since it is intended to be displayed on language maps and easily read by non-specialists. However, all phonemic distinctions are marked, and additional characters are introduced when necessary. In the case of Kordestan Province, phonemic symbols introduced from the wider field of Kurdish linguistics included ē, ɪ, r̰ and l̰. A subsequent critique of the research process by a workshop team in December 2016 at Allameh Tabataba'i University, Tehran, highlighted the value of also including a phonetic transcription of each data item. However, this was not part of the research process for Kordestan Province, which was carried out during the previous year.

provinces of Iran (see Section 1 "Introduction" above), this experience has turned out to be valuable in preparing Atlas team members for selection of research sites for extended language data collection, and for the actual linguistic structures they will encounter when carrying out this subsequent step.

In terms of the actual local place name data, which is available for each settlement on the Atlas website, we observed three patterns regarding correspondences between local names and the official Persian labels in the 2011 census data (ISC 2011). In most cases, local place names are identical or correspond in a systematic way, whether phonologically or lexically, to the official Persian labels:

Local name	Official name
Bijār	Bijār
Kāni Dırēzh	Cheshmeh Derāz
Kawpēch	Kowpich
Ōghaḷ	Owghal
Qurwa	Qorveh

In many other cases, there is still a clear resemblance, but the official label has been assigned with a similar – and often slightly longer – Persian term. Some of the differences can be attributed to the shortening of local names through a natural process of phonological reduction in the spoken language, but for other pairs of similar-sounding items the official label reflects a Persian folk etymology (that is, a semantic reinterpretation) of the local place name.

Local name	Official name
Bāyzāwā	Bāyzid Ābād
Jērāmina	Jeyrān Mangeh
Kōsawmar	Kows Anbar
Mirasām	Mir Hesām
Sına	Sanandaj

Finally, it is occasionally the case that a complete different official label has been applied to a given settlement, but the original name has been retained locally alongside or instead of the official label:

Local name	Official name
Biyakaṛa	Hoseyn Ābād
Kharka	Bahārestān
Khöḷına	Zafar Ābād
Māma Shā	Eslām Ābād
Say Ismāil	Hay'at Ābād

Although it is not the focus of the present study, even a cursory observation of this third set of place names reveals certain tendencies in the official naming and renaming of settlements, with a preference for labels that reflect the national language, culture, and official confession.

5 Language distribution and classification

In this section, which forms the core of the study, we investigate, analyze and describe patterns of language distribution and classification for Kordestan Province. After a presentation and discussion of these interconnected research questions, we provide an overview of language distribution of the province, and introduce a map which visualizes the results of our research. We then provide a detailed classification and discussion of varieties, addressing linguistic as well as sociolinguistic considerations. Extended discussion is devoted to the particularly complex status and internal classification of Central Kurdish and Southern Kurdish. After highlighting outstanding issues and questions in classification, we outline future directions for a refined classification of the languages of Kordestan Province.

5.1 Research questions for language distribution: Focus and limitations

For the topic of language distribution, which is a central theme in the Atlas, we limited our investigation to the following two questions (repeated from Section 3.3 above): 1) *What languages, and what subvarieties of these languages, are spoken as a mother tongue in this settlement?* 2) *In the case that more than one variety is spoken in the settlement, what is the estimated proportion of mother tongue speakers of each variety?*

As stated above, our objective in collecting these data has been to assemble a first coherent and detailed picture of language distribution for all of

Kordestan Province. While this is a worthwhile initiative in itself, it is indispensable to further research toward the Atlas' central purpose of providing a systematic investigation of key linguistic structures in the region as a whole, including linguistic characteristics of each dialect, as well as similarities and differences among dialects. Examination of the language distribution data has helped us to optimize our selection of sites for subsequent in-depth language data collection (sociolinguistic context, lexicon, phonology, morphosyntax, texts) across the province using the ALI questionnaire (`http://carleton.ca/iran/questionnaires`).

Multilingual proficiency in Persian and other languages, and other major sociolinguistic factors such as language change, shift and endangerment, are also essential in understanding the language situation as a whole. However, such factors are better suited to community-specific sociolinguistic studies such as those featured in Sheyholislami & Sharifi (2016) and Shahidi (2008) and Anonby & Yousefian (2011). For this reason, further sociolinguistic topics are reserved for systematic inquiry in the research sites where in-depth language data collection will take place using the Atlas questionnaire.

5.2 An overview of language distribution in Kordestan Province

As mentioned at the beginning of this article, Kordestan Province is often viewed as a linguistically homogeneous area, with Central Kurdish (or "Sōrāni"; but see Section 5.3 below for a discussion of this label) as the characteristic mother tongue for the province as a whole. While this depiction is legitimate in a very general sense, our comprehensive investigation of language distribution across the province highlights significant linguistic diversity of three types: the existence of several major language groupings; significant internal dialectal diversity in several of these languages; and important social factors correlated with language distribution.

The following map, which is the first systematic overview of language distribution in Kordestan Province – to the level of each settlement – summarizes the results of our research.

As the map shows, the major language groupings across the province are as follows:

[5] While the map printed here indicates only the language with the largest proportion of speakers in each settlement, the online map provides proportions for speakers for each language in settlements where more than one variety is spoken.

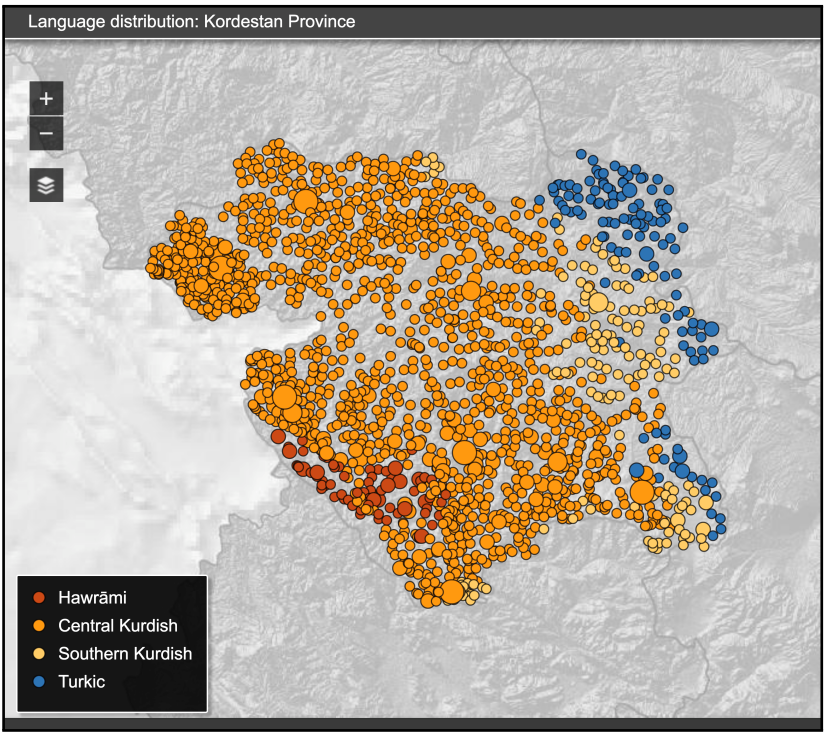

Figure 2: Overview of language distribution in Kordestan Province[5] (`http://iranatlas.net/module/language-distribution.kordestan`)

1. **Central Kurdish**, found across much of the province;

2. **Southern Kurdish**, found in pockets in several parts of the province: in the north-east and south-east corners of the province, in the regions of Bijār and Qorveh respectively; in Kāmyārān region, along the southern border with Kermanshah Province; and in a handful of villages near Saqqez in the north-west.

3. **Hawrāmi**, spoken in many settlements in the south-western part of the province, but also as a significant minority in the cities of Sanandaj and Marivān, and in the village of Qallā near Qorveh; and

4. **Turkic**, spoken in the outer areas of Bijār and Qorveh districts in the east.

There are also significant **Persian**- and **Aramaic**-speaking minorities in the cities of the province (see discussion in Section 5.3 below).

The distribution of Southern Kurdish, Hawrāmi and Turkic is fascinating, because each of these groups covers a significant area of the map. As was the case for Mohammadirad, a linguist from Kordestan who learned much about the language situation in the province over course of his research (see Section 3.3 above), the extent and nature of linguistic diversity can be surprising and informative for scholars as well as inhabitants of the province.

5.3 Classification and description of language varieties

Although each component of the present research was accompanied by a unique set of challenges, the classification and labelling of language varieties in Kordestan Province turned out to be one of the more intricate tasks, and required an additional phase of research and analysis after all other components were completed. The development of a coherent picture of language classification – itself a multi-faceted and exceedingly complex enterprise – for a vast, poorly-documented geographic area, necessitated an approach that was both innovative and flexible. At the same time, our integration of new insights and incorporation of more specific research questions over the course of the study, as described in this section below, meant that the research process was not identical for all locations.

To arrive at an overarching taxonomy of all language varieties in the province, we looked first of all at how people perceive and describe each language variety (see Anonby et al. 2016), with attention to several kinds of labels (some of which can overlap):

- **autoglottonyms**: labels that speakers use to refer to their own language variety;
- **heteroglottonyms**: language variety labels used by people who are not speakers of a given variety – whether speakers of related varieties, speakers of other languages in the region, or people outside the region;
- **labels applied by linguists**, which, strictly speaking, are a further type of heteroglottonym;

- **assessments**, by individuals (representing any of the three points above) with knowledge of the topic, as to whether two language varieties are the same or different; and

- **non-linguistic labels** applied to language varieties, such as geographic, ethnic or religious terms.

For Kurdish (Central Kurdish and Southern Kurdish together), Hawrami, Turkic, Aramaic and Persian, the designation of higher-level language groupings was straightforward, since speakers as well as scholars are easily able to distinguish each of these varieties from one another.[6]

On the other end of the question of classification, speakers of all the language varieties are typically comfortable referring to their specific language variety using the name of the exact settlement they come from.

Between these two ends of the spectrum, the labelling and classification of mid-level language varieties, which is the most valuable element in furthering what is known about the dialect situation in Kordestan Province, were harder to ascertain.[7]

In some cases, the different types of mid-level labels listed above correspond neatly, and in such cases we have chosen to accept these assessments, subject to further study on the linguistic structure of each variety, as a working hypothesis for their classification. In other cases, people are very con-

[6]The relationship between Hawrami on the one hand, versus Central and Southern Kurdish on the other, is multifaceted. There has been longstanding debate as to whether Hawrami is a Kurdish subvariety, closely related to Kurdish, or a historically distinct branch of Iranic, and the conclusions proposed by scholars (e.g., Soane 1921; Minorsky 1943; MacKenzie 1986, 1987, 2002; Kreyenbroek 1992; Leezenberg 1993; Hassanpour 1998; Haig & Öpengin 2014) differ according to whether social factors, synchronic structural similarities, or historical linguistic innovations are given precedence. Additional in-depth research, with careful attention to presuppositions about parameters for classification, is needed to clarify this question. In any case, this debate falls outside of the present study, since both Hawrami and Kurdish speakers consulted here consistently make primary reference refer to this variety as Hawrami or an equivalent label such as *māchō māchō* or *māchō zwān* (see the discussion on Hawrami in the later part of this section); this label therefore constitutes one of the higher-level language groupings for Kordestan Province.

[7]This situation is actually the opposite of what we observed in our dialectological work in Hormozgan Province (Anonby & Yousefian 2011; Mohebbi Bahmani et al. 2015; Anonby & Mohebbi Bahmani 2016; Nourzaei et al. 2015; Anonby 2016). There, mid-level language groupings (e.g., "Achomi", Keshmi, Bandari of Bandar Abbas, Minabi, Rudoni, Koroshi, Marzi Gal Bashkardi, etc.) are clearly defined and generally agreed-upon by speakers and others, including scholars. Conversely, in most cases, the designation of a "language" label for these varieties (e.g., "Persian", "Lārestāni", "Bandari", "Hormozgani", "Balochi", etc.), and the grouping of the varieties under these common labels, is problematic for speakers and scholars alike).

scious of a difference between varieties but do not have labels to distinguish them, and in such cases (which are clearly indicated), we have proposed labels for these groupings. Further, there are cases where linguists, and especially the field linguist Mohammadirad, have observed systematic structural differences between varieties, and this informs our proposal for how varieties can best be classified. Examples of each of these situations are provided in the ensuing discussion.

With these factors as a backdrop, the following subsections provide further discussion and detailed internal classifications of the language varieties of Kordestan Province. All language classifications are abstractions, and are necessarily based on a consideration of finite sets of linguistic and sociolinguistic factors. In many cases, even when all factors are adequately investigated and controlled, there is still limited consensus regarding any language classification (see Anonby et al. 2016 for further discussion).

In the end, we submit the classification developed here as working model that will facilitate refinement of scholarly understanding of the language situation in Kordestan Province, through dialogue and the collection of language data that speak directly to relationships among varieties.

5.3.1 Defining Central Kurdish and Southern Kurdish

As observed in the literature (Haig & Öpengin 2014: 103; Sheyholislami 2015: 30) and confirmed by Mohammadirad during fieldwork for the present study, speakers of Central Kurdish and Southern Kurdish consistently make primary reference to their language simply as *kurdi* 'Kurdish'.

When consulted about what kind of Kurdish they speak, respondents generally refer first to a very local variety (Kurdish of a given village), or a mid-level variety such as "Ardalāni" or "Garūsi" (for other examples, see the internal classifications of Central Kurdish and Southern Kurdish below). High-level groupings equivalent to Central Kurdish and Southern Kurdish are rarely used as autoglottonyms.

However, and as reflected in the assessments of many scholars (e.g., Paul 1998; Fattah 2000; Korn 2003; Windfuhr 2009; Mohammadirad, field notes 2016), Central Kurdish and Southern Kurdish are recognized as important high-level groupings by speakers. This fact emerged with speakers in the course field research: in response to the question, "Are there some main groups of Kurdish in Kordestan Province?", respondents generally identified *kursāni* 'Kordestani' and *kirmāshāni* 'Kermanshahi' as two high-level groupings. Of course, this labelling is problematic since, as clearly demonstrated in

our study and elsewhere, neither variety is by any means limited to these respective provinces. There are also Kurdish-speaking groups at the periphery that view *kursāni* as referring only to the dialects in the centre of the province. In districts where both Central Kurdish and Southern Kurdish are spoken, the labelling of the two groups was more nuanced and specific: in Kāmyārān district, for example, Central Kurdish is referred to simply as *kurdi*, but Southern Kurdish is known as *arā zwān* ("*arā*" language). This label refers to the frequently occurring Southern Kurdish morpheme *arā* 'why, for', which is conspicuously different from its Central Kurdish counterpart *bō*.

Among scholars, and especially those of Western origin (e.g., MacKenzie 1961, 1962), "Sōrāni" has been used as a general term for Central Kurdish, in contradistinction to the term "Kurmanji" used to refer to Northern Kurdish. However, as will be evident from the discussion of Central Kurdish varieties below, the label "Sōrāni" is not typically used by Central Kurdish speakers of Kordestan Province to refer to Central Kurdish as a whole. Because of this (and along with considerations relevant to other parts of the language area), scholars are increasingly referring to this variety as Central Kurdish. In contrast, as the recognition of the Central Kurdish variety becomes more prominent in popular Kurdish discourse, MacKenzie's original label of "Sōrāni" is increasingly used by speakers in parts of the Central Kurdish language area which did not formerly use this label (Hassanpour 2012; Sheyholislami 2012; Mohammadirad, field notes 2016).

5.3.2 Internal classification of Central Kurdish

Based on the results of our initial field research, we propose the classification of Central Kurdish in Kordestan Province into the following five subvarieties:

- **Sōrāni**, as defined in this study,[8] is spoken in northern and western districts (Persian: *shahrestān*) of Kordestan Province: Saqqez, Bāneh, Marivān, and parts of Sarv Ābād. The label for this subvariety, which is used by speakers, comes from the former principality of Sōrān (Hassanpour 1992; McDowall 2004), which is located in the north-eastern part of present-day Kurdistan Governorate in Iraq, close to the Iranian border. Even within Sōrāni, there are some geographically or socially

[8] While the term "Sōrāni" has been used in the literature to refer to all Central Kurdish subvarieties, speakers in Kordestan Province see it as a subset of Central Kurdish (see "Defining Central Kurdish and Southern Kurdish" above).

defined dialect groupings (Tilakōyi, Sarshiwi, Fayzuḷābagi,[9] etc.) that speakers use for their own variety, but we have been unable to observe or posit such groupings across all of the Sōrāni-speaking areas. A number of the city/district-inspired labels (Mariwāni, Bānayi, Saqezi, etc.) are used by Central Kurdish speakers from elsewhere in the province, but not by the speakers who are from the given area. Central Kurdish speakers in Divān Darreh district also use the label "Sōrāni" for their variety, but Mohammadirad has observed through fieldwork that the dialect spoken in Divān Darreh district appears to be transitional between Ardalāni and the Sōrāni dialects to the west, and that – subject to further study – it may be better classified with Ardalāni if structural linguistic considerations are given precedence.

- **Ardalāni** is centred in Sanandaj city (K. *sına*) and spoken throughout Sanandaj district. The name for this subvariety, which is used by speakers, comes from the former Ardalān Principality, which was the last principality to be dismantled by Ottoman and Persian empires (Hassanpour 1992). The subvariety is referred to as *sınayi* 'Sanandaji' by people from elsewhere, but within the Ardalānī area, the application of the term *sınayi* tends to be limited to the dialect of Sanandaj city, and its conservative form in particular. As mentioned in the previous paragraph, the Central Kurdish subvariety spoken in Divān Darreh district appears to be transitional between Ardalāni and the Sōrāni dialects to the west, and if fact may pattern more closely with Ardalāni in its structure.

- **Laylākhi** is spoken in Deh Golān district and the western side of Qorveh district. The name of this subvariety, which is used by speakers, was reflected in the name of the former district which included both current districts. Speakers also refer to Laylākhi as "Gōrāni" to distinguish it from other Central Kurdish varieties, but it is not related to the Gōrāni linguistic group of which Hawrāmi is a part (see Mahmoudveysi et al. 2012 for the geographic dispersion of this latter group). As is the case for varieties within Sōrāni, speakers point to internal divisions

[9] The first two varieties mentioned here (Tilakōyi and Sarshiwi) are geographically defined, and take their names from the main cities of the dialect area, but the Fayzuḷābagi variety has a social component as well: it is the "original" high-prestige Sōrāni variety spoken in and around the city of Saqqez, and is considered as different from the "new" Sōrāni that is increasingly found there.

within Laylākhi using geographic and social labels (Qurwai, Shēkh Esmāili, etc.), but according to Mohammadirad (field notes 2016), the linguistic basis of these divisions is difficult to establish and deserves further investigation.

- **Bijār Central Kurdish** is spoken by a minority of people in Bijār city, and in many villages in the western part of Bijār district. This label is not used by speakers, who simply refer to their language as *kurdi* 'Kurdish', in distinction to the Southern Kurdish varieties that dominate the district, which they refer to as *bijāri* or *garūsi* (see "Garūsi" under "Southern Kurdish" below).

- **Kāmyārān Central Kurdish** is spoken by a majority of people in Kāmyārān city and the district of Kāmyārān as a whole. This label is not used by speakers, who simply refer to their language as *kurdi* 'Kurdish', in distinction to the Southern Kurdish varieties spoken in the district and in Kermanshah Province to the south, which they refer to as *arā zwan* "'why, for" language'.

The classification of subvarieties of Central Kurdish, as summarized here, reveals a set of five more- or less-well-defined subvarieties that can serve as a useful starting point for further dialectological research. We expect significant variation within each of the five subvarieties and, as treated explicitly for Central Kurdish of Divān Darreh, there are almost certainly transitional areas between each of the subvarieties. The nature of intra- and inter-dialect variation can be investigated further when language data is collected from major dialect centres and transitional areas at the edges of each subvariety.

5.3.3 Internal classification of Southern Kurdish

Southern Kurdish varieties in Kordestan Province are easier to classify than Central Kurdish varieties for several reasons: they are geographically discrete, separated by Central Kurdish-speaking areas; for most of the varieties, there are clear labels that distinguish them from Central Kurdish; and speakers' assessments of their own dialects match those of other groups. The four main Southern Kurdish subvarieties in the province are as follows:

- **Garūsi** is spoken in the city of Bijār, along with most villages in central and southern portions of Bijār district. This variety is named after the former principality of Garrus (Hassanpour 1992).

- **Chardāwri** (which Fattah 2000 refers to as Chahār Dawli) is spoken by a minority of the population in Qorveh city, and in most of the towns and villages to the south-east, up to the borders of Hamadan and Kermanshah Provinces. The name of this variety is derived from the Chardāvol district of Ilam Province, where speakers situate their origins. Along with this label, speakers also refer to their variety as *kulyāyi*, since it is similar to Kulyāyi varieties of Southern Kurdish in neighbouring areas across the border in Kermanshah Province.
- **Kāmyārān Southern Kurdish** is spoken by a significant minority of people in Kāmyārān city, and in almost twenty villages in the south-east corner of Kāmyārān district, toward the border with Kermanshah Province.
- **Kalhuri** is spoken in a handful of villages at the far north-east corner of Saqqez district, on the border with West Azerbaijan Province. The speakers of this subvariety come originally from the larger Kalhuri-speaking region that spans the border of Kermanshah and Ilam Provinces to the south.

There is also a handful of mixed Central and Southern Kurdish-speaking villages (or perhaps a transitional variety between them) found at the south end of Deh Golān district, on the border of Kermanshah Province. We do not yet have any detailed information about the language spoken in these villages, and until a fuller understanding of the dialect situation in the adjacent areas of Kermanshah Province becomes available (ongoing work is published in Fattahi et al. 2018), it will be difficult to provide an assessment or classification.

5.3.4 Hawrami

Hawrami (autoglottonym: *hōrāmī*), known as *māchō māchō* or *māchō zwān* in Kurdish,[10] is represented by two major subvarieties in the south-west portion of Kordestan Province: **Takht Hawrami** in the south-west corner, and **Zhāwarū Hawrami**, toward the interior of the province. While both varieties are centred in Sarv Ābād district, there is a significant minority of Takht Hawrami speakers in Marivān city; and Zhāwarū Hawrami-speaking villages

[10]The label is an imitation of the Hawrami term *māchō* 'he/she said', a common structure that differs from its Central Kurdish counterparts *daḷē* or *ayzhē*, depending on the region (Mohammadirad, field notes 2016).

extend into the districts of Sanandaj and Kāmyārān. In south-east Kordestan, the Hawrami-speaking village of Qallā, now administratively part of Qorveh city, refers to its own variety as **Qaḷāyi**; the relation between this and other Hawrami varieties has yet to be studied.

5.3.5 Turkic

From our initial survey, there are three main subvarieties of Turkic in Kordestan Province:

- **Shāhsevan Turkic** is spoken in part of Bijār district. This variety is associated with the traditionally migratory Shāhsevan ethnic group (Tapper 2010), which extends to several other provinces of Iran.
- So-called **"Tāt" Turkic**, as it is referred to by its speakers, is spoken in other parts of Bijār district. Considering its label, it is possible that the people who speak this subvariety were originally speakers of the Northwestern Iranic language Tāti.
- **Ghürva** (or Qorveh) **Turkic**, a dialect group known to its speakers simply as "Torki", is spoken by a minority of people in the city of Qorveh, and in about twenty towns and villages on the eastern side of the district, up to and along the border with Hamadan Province. Speakers often refer to their language according to the individual clans they belong to (Notarki, Khodābandelu, Bāghluja, etc.); however, we have not observed any linguistic basis for this type of further subdivision of Ghürva Turkic.

None of the Turkic varieties within Kordestan Province has ever been described in the literature, although the existence of Shahsavān Turkic is at least known from other areas of Iran. Consequently, there is a great need for further study of this topic.

5.3.6 Persian

While Persian is not indigenous to Kordestan, it is spoken by immigrants to some of the larger cities of the province. In addition, as observed during Mohammadirad's research for this study, Persian is emerging as a mother tongue in some of the areas of the province, as parents teach it to their children as a first language at home. Although this situation is observable in

Sanandaj and other larger cities of the province (as is the case in cities throughout Iran), the trend is most advanced in the eastern cities of Bijār and Qorveh, where among the urban Turkic communities it dates to the era before the Islamic Revolution. In contrast to the rest of the province, the cities of Bijār and Qorveh are predominantly Shi'a, and this is correlated with a favourable disposition toward Persian as a national language. Alongside existing positive attitudes, the cohabitation of Turkic and Kurdish and communities within the cities has promoted the use of Persian as a language of wider communication. Specifically, an increasing incidence of linguistically mixed marriages is a predictive factor in which homes children learn Persian as a mother tongue; however, an increasing number of parents in non-mixed Turkic and Kurdish homes are also teaching their children Persian (Mohammadirad, field notes 2016).

5.3.7 Aramaic

Before the Islamic Revolution, there were significant populations of Aramaic speakers (autoglottonyms: *ārāmāyā* 'Aramaic', *lishāna nōshan* 'our language') in Sanandaj as well as Bijār. As we learned during fieldwork among diaspora speakers of the language in New York City (Hoberman, Borjian and Anonby, field notes 2014), there is still a remnant of the former language communities in each city, with an even smaller subset of the communities – mostly older individuals – by whom Aramaic is still spoken. The Aramaic populations of these cities, along with speakers in the diaspora, share a single variety of North-Eastern Neo-Aramaic (Hoberman, Borjian and Anonby, field notes 2014; Geoffrey Kahn, pers. comm. 2017; see also Rosenthal 1986 and Windfuhr 2006).

6 Conclusion

In this paper, we have provided an account of the research process and results for Kordestan Province within the *Atlas of the Languages of Iran* research programme. In order to produce this Atlas module, a large team of scholars has carried out research of various types: collection and processing of existing geographic and demographic data for Kordestan Province; construction of a linguistic bibliography; compilation of local names for all settlements in the province; assessments of language and dialect distribution for each settlement; and publication of an open-access online map that embraces all of these elements, making them available to scholars and popular audiences.

The key result of this study is a first comprehensive picture of language distribution in Kordestan Province, with detail provided to the level of each settlement. This study shows, in contrast to prevalent conceptions, that Kordestan Province is linguistically diverse, with six important language groups represented: Central Kurdish, Southern Kurdish, Hawrami, Turkic, Persian and Aramaic. This diversity is also reflected by internal dialectal variety within the major groups. In the case of Central Kurdish and Southern Kurdish, we have provided a thorough initial classification of major subvarieties in the province, taking into consideration a range of linguistic and extralinguistic factors. We have also proposed initial classifications for the Hawrami and Turkic varieties spoken in the region, though many questions remain.

The results of the research presented in this article are not a final delineation of all aspects of the linguistic situation. Rather, along with the accompanying open-access resources published in the Atlas, they are intended as a catalyst and guide to further inquiry. In one such application of the results, this study is enabling the next phase of Atlas research on Kordestan Province: the selection and implementation of language data collection for using the Atlas questionnaire. In this way, and in conjunction with other research initiatives carried out by other scholars in the field, our understanding of the language situation in Kordestan Province will be progressively refined.

References

Aliakbari, Mohammad & Gheitasi, Mojtaba & Anonby, Erik. 2014. On language distribution in Ilam Province, Iran. *Iranian studies* 48(6). 835–850. (`https://doi.org/10.1080/00210862.2014.913423`).

Anonby, Erik. Forthcoming. Phonological variation in Kurdish. In Matras, Y. & Haig, G. & Öpengin, E. (eds.), *Variation and typology in the dialects of Kurdish*. London: Palgrave Macmillan.

Anonby, Erik. 2015. Mapping Iran's languages: Situation and prospects. *6th International Conference of Iranian Linguistics (ICIL6)*, Ilia State University, Tbilisi, Georgia.

Anonby, Erik. 2016. The Keshmi (Qeshmi) dialect of Hormozgan Province, Iran: A first account. *Studia iranica* 44(2). 165–206.

Anonby, Erik & Mohebbi Bahmani, Hassan. 2016. Shipwrecked and landlocked: Kholosi, an Indo-Aryan language in south-west Iran. *Studia iranica* 43 [supplement volume]. 1–23.

Anonby, Erik & Murasugi, Kumiko & Domínguez, Martín. In press. Mapping language and land with the *Nunaliit Atlas Framework*: Past, present and future. *Endangered languages and the land: Mapping landscapes of multilingualism. Proceedings of the 22nd Conference of the Foundation for Endangered Languages*, 23–25 August, 2018, Reykjavik, Iceland. 56–63.

Anonby, Erik & Sabethemmatabadi, Parisa. In press. Representation of complementary user perspectives in a language atlas. In Taylor, D. R. F. (ed.), *Further developments in the theory and practice of cybercartography: International dimensions and language mapping*. Amsterdam: Elsevier.

Anonby, Erik & Sabethemmatabadi, Parisa & Hayes, Amos. 2016. Reconciling contradictory perspectives of language identity in the Atlas of the Languages of Iran. Paper presented at the *Cognitive Science Colloquium Series*, Carleton University, February 25, 2016.

Anonby, Erik & Taheri-Ardali, Mortaza. 2018. Bakhtiari. In Khan, G. & Haig, G. (eds.), *Language contact and language change in west Asia*, 445–480. Berlin: De Gruyter.

Anonby, Erik & Taheri-Ardali, Mortaza, et al. (eds.). 2015–2019. *Atlas of the Languages of Iran*. Ottawa: Geomatics & Cartographic Research Centre, Carleton University. (`http://iranatlas.net/`) (Accessed 2019-06-15).

Anonby, Erik & Taheri-Ardali, Mortaza & Hayes, Amos. 2019. The Atlas of the Languages of Iran (ALI): A research overview. *Iranian studies* 52(1–2). 1–32.

Anonby, Erik & Yousefian, Pakzad. 2011. *Adaptive multilinguals: A study of language on Larak Island*. Uppsala: Acta Universitatis Upsaliensis.

Atlas Narodov Mira 1964 = Bruk, S. I. and Apenchenko, V. S. (eds.). 1964. *Atlas Narodov Mira*. Moscow: Glavnoye Upravleniye Geodezii i Kartografii.

Compendium 1989 = Schmitt, R. (ed.). 1989. *Compendium linguarum iranicarum*. Wiesbaden: Reichert.

Fattah, Ismail Kamāndār. 2000. *Les dialectes kurdes méridionaux: Étude linguistique et dialectologique* (Acta Iranica 37). Leuven: Peeters.

Fattahi, Mehdi et al. 2018. Kermanshah Province. In Anonby, E. & Taheri-Ardali, M., et al. (eds.), 2015–2019. *Atlas of the Languages of Iran*. Ottawa: Geomatics & Cartographic Research Centre, Carleton University. (`http://iranatlas.net/module/language-distribution.kermanshah`) (Accessed 2019-07-30).

GCRC (Geomatics and Cartographic Research Centre). 2013. *Nunaliit Atlas Framework*. Ottawa: GCRC, Carleton University. (`http://nunaliit.org`) (Accessed 2019-07-28).

Haig, Geoffrey & Öpengin, Ergin. 2014. Kurdish: A critical research overview. *Kurdish studies* 2(2) [Special issue: *Kurdish linguistics: Variation and change*]. 99–122.

Hasanzādeh, Behrooz. 1999. *A study of the verb system in the Kurdish language (Sanandaji dialect) and comparison with ancient languages of Iran*. Tehran: Islamic University, Olum va Tahqiqat Branch MA thesis.

Hassanpour, Amir. 1992. *Nationalism and language in Kurdistan 1918–1985*. San Francisco: Mellen Research University Press.

Hassanpour, Amir. 1998. The identity of Hewrami speakers: Reflections on the theory and ideology of comparative philology. In Soltani, A. (ed.), *Anthology of Gorani Kurdish poetry, compiled by A. M. Mardoukhi (1739–1797)*, 35–49. London: Soane Trust for Kurdistan.

Hassanpour, Amir. 2012. The indivisibility of the nation and its linguistic divisions. *International journal of the sociology of language* 217. 49–73.

Irancarto 2012 = Hourcade, Bernard et al. 2012. *Irancarto*. Paris/Tehran: CNRS/University of Tehran. (`http://www.irancarto.cnrs.fr`) (Accessed 2019-07-12).

ISC (Iran Statistics Center). 2011. *Public census of population and settlement 2011*. Tehran: Ministry of the Interior, Iran Statistics Center. (`https://www.amar.org.ir`) (Accessed 2019-07-31).

ISC (Iran Statistics Center). 2016. *Public census of population and settlement 2016*. Tehran: Ministry of the Interior, Iran Statistics Center. (`https://www.amar.org.ir`) (Accessed 2019-07-31).

Izady, Michael Mehrdad. 1998. *Linguistic composition of Kurdistan*. Paris: Institut kurde de Paris. (`http://www.institutkurde.org/images/cartes_and_maps/linguistic_composition.jpg`) (Accessed 2019-07-30).

Izady, Michael Mehrdad. 2006–13. *Linguistic composition of Iran*. (`http://gulf2000.columbia.edu/images/maps/Iran_Languages_lg.png`) (Accessed 2019-07-31).

Khādemi, Shirin. 2002. *National Project of Iranian Dialectology: A comparative study of lexicon and basic sentences in 40 villages of Kordestan and Kermanshah Provinces*. Tehran: Islamic Azad University (Central Tehran Branch) MA thesis.

Khaliqi, Shahla. 2001. *A study of the duality of phonological analysis in the Saqezi dialect of Kurdish*. Tehran: University of Tehran MA thesis.

Kordzafarānlu Kāmbuziā, Āliyeh & Sajjadi, Seyyed Mahdi. 2013. Syllable structure in Hawrami (Hawraman Takht dialect). *Journal of Western Iranian languages and dialects* 1(2). 57–78.

Korn, Agnes. 2003. Balochi and the concept of North-West Iranian. In Jahani, C. & Korn, A. (eds.), *The Baloch and their neighbours: Ethnic and linguistic contact in Balochistan in historical and modern times*, 49–60. Wiesbaden: Reichert.

Kreyenbroek, Philip. 1992. On the Kurdish language. In Kreyenbroek, P. & Speril, S. (eds.), *The Kurds: A contemporary overview*, 68–83. London: Routledge.

Leezenberg, Michiel. 1993. *Gorani influence on Central Kurdish: Substratum or prestige borrowing?* Amsterdam: Institute for Logic, Language & Computation (ILLC), University of Amsterdam.

MacKenzie, David Neil. 1961. *Kurdish dialect studies*. Vol. 1. London: Oxford University Press.

MacKenzie, David Neil. 1962. *Kurdish dialect studies*. Vol. 2. London: Oxford University Press.

MacKenzie, David Neil. 1986. Kurds, Kurdistan: Language. In Bosworth, C. E. (ed.), *Encyclopaedia of Islam*, vol. 5, 479–480. Leiden: Brill.

MacKenzie, David Neil. 1987. Avromani. In *Encyclopaedia iranica*, vol. III.1, 111–112. New York: Columbia University. (`http://www.iranicaonline.org/articles/avromani-the-dialect-of-avroman-q`) (Accessed 2019-07-31).

MacKenzie, David Neil. 2002. Gurāni. In *Encyclopaedia iranica*, vol. XI.4, 401–403. New York: Columbia University. (`http://www.iranicaonline.org/articles/gurani`) (Accessed 2019-07-31).

Mahmoudveysi, Parvin & Bailey, Denise & Paul, Ludwig & Haig, Geoffrey. 2012. *The Gorani language of Gawraju (Gawrajuyi), a village of west Iran: Grammar, texts and lexicon*. Wiesbaden: Reichert.

Matras, Yaron & Koontz-Garboden, Andrew. 2017. *The dialects of Kurdish*. Manchester: University of Manchester. (`http://kurdish.humanities.manchester.ac.uk`) (Accessed 2019-07-31).

McDowall, David. 2004. *A modern history of the Kurds*. London: I. B. Tauris.

Minorsky, Vladimir. 1943. The Guran. *Bulletin of the School of Oriental and African studies* 11(1). 75–103.

Mohammadi, Sho'leh. 2002. *A description and analysis of the phonetic system of the Kurdish language (Sanandaji dialect) based on Generative Phonology and Autosegmental Phonology*. Tehran: Allameh Tabatabai University MA thesis.

Mohebbi Bahmani, Hassan et al. 2015. Hormozgan Province. In Anonby, Erik & Taheri-Ardali, Mortaza, et al. (eds.), 2015–2019. *Atlas of the Languages of Iran*. Ottawa: Geomatics & Cartographic Research Centre, Carleton Uni-

versity. (http://iranatlas.net/module/language-distribution.hormozgan) (Accessed 2019-07-12).

NCC (National Cartographic Center). 2014. *Iran: Administrative districts and boundaries*. Tehran: Ministry of the Interior, National Cartographic Center. (https://www.iscgm.org/gmd). Available online until August 2016.

NCC (National Cartographic Center). 2016. *National database of geographic names of Iran*. Tehran: Ministry of the Interior, National Cartographic Center. (http://gndb.ncc.org.ir) (Accessed 2019-07-30).

Nourzaei, Maryam & Jahani, Carina & Anonby, Erik & Ahangar, Abbas Ali. 2015. *Koroshi: A corpus-based grammatical description*. Uppsala: Acta Universitatis Upsaliensis.

Paul, Ludwig. 1998. The position of Zazaki among West Iranian languages. In Sims-Williams, N. (ed.), *Proceedings of the Third European Conference of Iranian Studies*, 11–15 September 1995, Cambridge. Part I: *Old and Middle Iranian studies*, 163–177. Wiesbaden: Reichert.

Rezāi, Vāli. 1996. *A linguistic description of the Sanandaji dialect of Kurdish*. Esfahan: University of Esfahan MA thesis.

Roostanet 2016 = Rural Network of Iran. 2016. *Roostanet*. Tehran: Rural Network of Iran. (http://map.roostanet.com) (Accessed 2019-07-30).

Rosenthal, Franz. 1986. Aramaic. In *Encyclopaedia iranica*, vol. II.3, 250–261. New York: Columbia University. (http://www.iranicaonline.org/articles/aramaic-#pt1) (Accessed 2019-07-12).

Sajjadi, Seyyed Mahdi & Kordzafarānlu Kāmbuziā, Āliyeh. 2014. Consonantal phonological processes in words common to the Standard Persian language and the Hawrami language (Hawramān Takht dialect). *Journal of Western Iranian languages and dialects* 1(4). 73–100.

Shahidi, Minoo. 2008. *A sociolinguistic study of language shift in Mazandaran*. Uppsala: Acta Universitatis Upsaliensis.

Sheyholislami, Jaffer. 2012. Kurdish in Iran: A case of restricted and controlled tolerance. *International journal of the sociology of language* 217. 19–47.

Sheyholislami, Jaffer. 2015. Language varieties of the Kurds. In Taucher, W. & Vogl, M. & Webinger, P. (eds.), *The Kurds: History, religion, language, politics*, 30–51. Vienna: Austrian Federal Ministry of the Interior.

Sheyholislami, Jaffer & Sharifi, Amir. 2016. It is the hardest to keep: Kurdish as a heritage language in the United States. *International journal of the sociology of language* 237. 75–98.

Soane, Ely B. 1921. A short anthology of Guran poetry. *Journal of the Royal Asiatic Society* 1. 57–81.

Sohrābi, Rezā & Serish Ābādi, Ja'far. 2009. *Farhang-e Garus*. Erbil: Aras.

SRTM 2014 = NASA (North American Space Agency). 2014. *Shuttle radar topography mission (SRTM)*. Pasadena: California Institute of Technology Jet Propulsion Laboratory. (http://www.jpl.nasa.gov/srtm) (Accessed 2019-07-30).

Taheri-Ardali, Mortaza et al. 2015. Chahar Mahal va Bakhtiari Province. In Anonby, Erik & Taheri-Ardali, Mortaza, et al. (eds.), 2015–2019. *Atlas of the Languages of Iran*. Ottawa: Geomatics & Cartographic Research Centre, Carleton University. (http://iranatlas.net/module/language-distribution.chahar_mahal_va_bakhtiari) (Accessed 2019-08-01).

Taheri-Ardali, Mortaza & Anonby, Erik. 2019. Some salient features of the Bakhtiari language in Chahar Mahal va Bakhtiari Province. In *Proceedings of the National Conference on Bakhtiari Language, Culture and Literature*, 9 May 2018, Payam-e Noor University of Shahr-e Kord, 993–1001. Tehran, Iran: Payam-e Noor University. (http://conference.pnu.ac.ir/ChaharMahal-bakhtiari/post.aspx?id=1312) (Accessed 2019-07-30).

Tapper, Richard. 2010. Shahsevan. In *Encyclopaedia iranica online*. New York: Columbia University. (http://www.iranicaonline.org/articles/shahsevan) (Accessed 2019-07-28).

TAVO 1988 = Orywal, E. (ed.). 1988. Karte 10, Vorderer Orient: Sprachen und Dialekte. *Tübinger Atlas des Vorderen Orients (TAVO)*. Vol. 8 (Series A). Wiesbaden: Reichert.

Teymuri, Robab. 1998. *A comparison of verb structure in the Garusi and Mahābādi dialects*. Tehran: University of Tehran MA thesis.

Windfuhr, Gernot. 2006. Iran vii. Non-Iranian languages. In *Encyclopaedia iranica*, vol. XIII.4, chap. 10. Aramaic, 404–410. New York: Columbia University. (http://www.iranicaonline.org/articles/iran-vii10-aramaic) (Accessed 2019-07-31).

Windfuhr, Gernot (ed.). 2009. *The Iranian languages*. London/New York: Routledge.

3
Pharyngeals in Kurmanji Kurdish: A reanalysis of their source and status

Daniel Barry

Abstract: A noteworthy feature of a number of Western Iranian languages, including Kurmanji Kurdish, is the presence of contrastive pharyngeal sounds in inherited vocabulary. These pharyngeals are considered by many linguists working on Kurdish to be the result of contact with Arabic, coming into the language through Arabic loan vocabulary (Haig & Matras 2002). The Arabic contact source of these sounds seems likely, particularly given the fact that most of the Western Iranian languages which contain pharyngeals are in contact with Arabic at present or historically. However, as I demonstrate, the distribution of the majority of contrastive pharyngeals in inherited Iranian vocabulary in Kurmanji does not suggest a mere surface imitation of Arabic vocabulary, but a Kurmanji-internal phonological process modulated by familiarity with the phonetics of Arabic pharyngeals. A newly-identified sound pattern presented here is the association of what are arguably pharyngealized vowel phonemes in Kurmanji with pre-existing labial consonants and constraints determined by Kurmanji phonotactics. Following Blevins' (2017) model of "perceptual magnets", this effect is held to have emerged on a model of Arabic pharyngeals as external "perceptual magnets" for native speakers of Kurdish who had extensive exposure to Arabic sound patterns.

1 Introduction

Kurmanji Kurdish is an Indo-European language belonging to the Western Iranian branch of the Indo-Iranian family. It is spoken natively in a region known locally as Kurdistan, which is divided between the states of Turkey, Iran, Iraq, and Syria. Kurdistan is a region which contains several Kurdish varieties that are considered "dialects" of Kurdish much as the various splits in Chinese are identified as "dialects", as well as two other Iranian languages (Hawrami and Zaza), which, for cultural reasons, are often referred to as "Kurdish", although, in a linguistic sense, they are less-closely related, and the exact nature of their common ancestry is less clear (Haig & Öpengin 2014: 111). But, in any event, the exact pedigree of "Kurdish" by any definition is not uncontroversial among linguists (Haig & Öpengin 2014; Paul 2008). All of these languages have come into some degree of contact with local Arabic and Turkic varieties.

Together with the closely-related Sorani Kurdish, Kurmanji Kurdish is one of the westernmost Iranian languages, and is marked by a more prolonged direct contact with Arabic compared to most other Iranian varieties. It is not surprising then that the presence of pharyngeal sounds in both inherited and loan vocabulary in Kurdish has been ascribed to Arabic contact (Haig & Matras 2002), although others have ascribed it to pre-Islamic contact with Aramaic (Hoberman 1985: 229). Regardless of the source, the introduction of new phonemes into a language based on contact with an unrelated language, but extending into the inherited vocabulary, is a phenomenon with important theoretical implications for phonological theory. The case of Kurmanji pharyngeals in particular is important due to the cross-linguistic rarity of pharyngeal sounds and their geographical restriction to a small number of areas (Blevins 2004: 197). The difficulty of accounting for these sounds in the inherited vocabulary lies in the apparent lack of any consonantal or vocalic source in Proto-Indo-European (PIE), or any apparent correspondence in other Iranian languages (e.g. Persian, see Table 1).

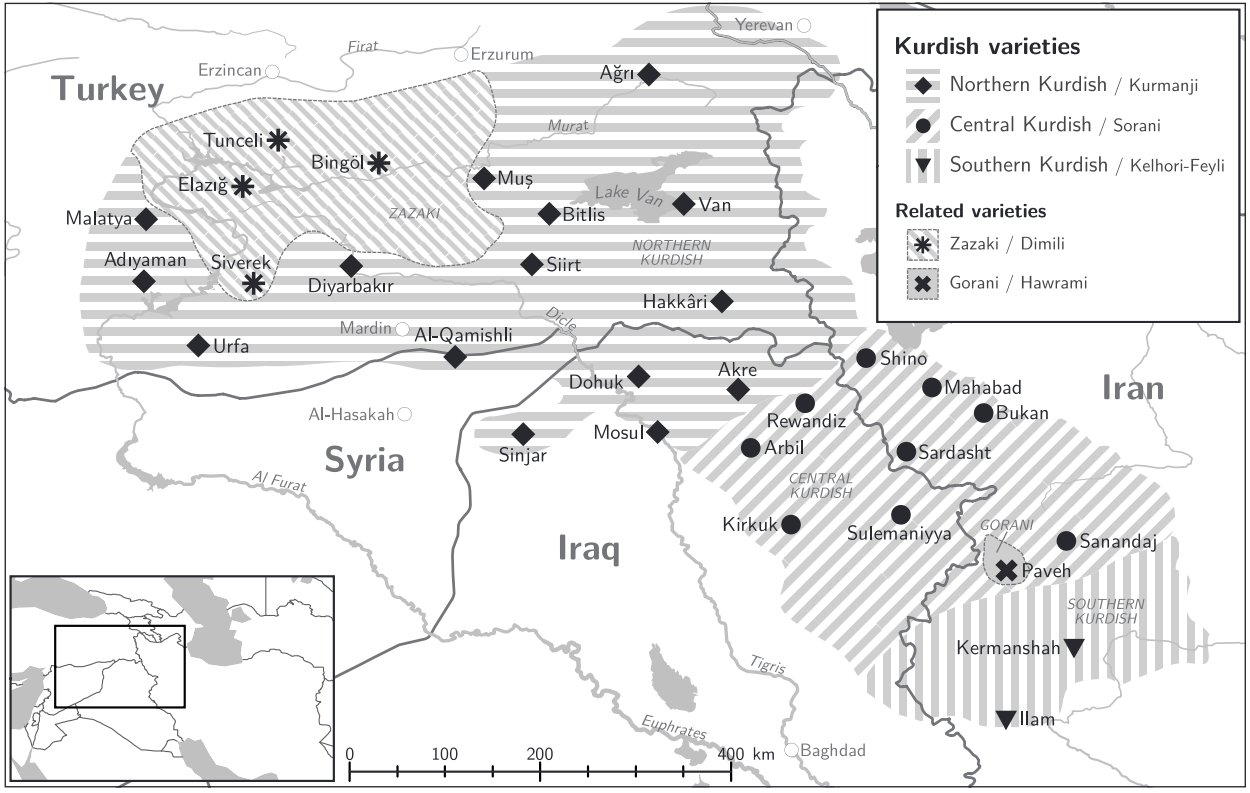

Figure 1: Linguistic map of Kurdistan (Haig 2018: 107)

Table 1: Selected Kurmanji pharyngeals with Persian cognates and Proto-Indo-European roots

#	Proto-Indo-European (Fortson IV 2010, unless noted)	PIE Gloss	Persian	Kurmanji Kurdish	Iranian Gloss (if different from PIE Gloss)
(a)	*n(e)b^h-ro- (Pokorny 2007)	'cloud'	[æbɾ]	[ʕæwɾ]	
(b)	*h_2e$\acute{k}$-men-	'stone'	[ɒːsmɒːn]	[ʕæzmɑn]	'firmament'
(c)	*ǵneh$_{3-}$	'to know'	[dɒːn]-	[zˁæn]-	
(d)	*septm̥	'seven'	[hæft]	[ħæft]	
(e)	*me$\acute{g}$h$_2$-	'great'	[meh]	[mæzˁɪn]	
(f)	*pek^w-	'to bake'	[poxt]	[pæħt] ~ [pæʕt]	'baked' (past stem)
(g)	*tep- (Pokorny 2007)	'to be warm'	[tɒːb]	[tæʕv]	Persian 'heat', Kurmanji Kurdish 'sun'
(h)	*sauso-	'dry'	[xoɟk]	[ħɪɟk]	'dry, arid', also 'solid' in Kurmanji Kurdish
(i)	*o$\acute{k}$-tō(u)	'eight'	[hæɟt]	[ħæɟt]	

In Table 1, there is no obvious PIE sound that corresponds to pharyngeals in Kurmanji: syllabic nasals such as the one in *n(e)b^{h}-ro- became *a by the stage of Proto-Indo-Iranian (Fortson IV 2010: 204), thus in Kurmanji Kurdish we expect [æwɾ], much like the Persian [æbɾ] ('cloud'). While laryngeals are held to have left traces in Iranian languages (Fortson IV 2010: 228), their preservation as initial consonants within the family is not documented or posited, making items like [ʕæzmɑn] ('firmament') anomalous versions of predicted forms like [ɑzmɑn]. Pharyngealized obstruents of the type found in Arabic also exist in some dialects of the language, posing similar problems based on reconstructed forms. In Sêrt and Bidlîs dialects, some examples of this in inherited vocabulary include [zˤæn]- and [mæzˤɪn], whose pharyngealized realization of Proto-Iranian *z- is straightforwardly from PIE *ǵ. PIE *ǵ is not regularly realized in these dialects with [zˤ] (cf. ǵenh$_1$, 'to beget', invariably non-pharyngeal, [zɑn] or [zɑjin]). PIE *s- is continued as /h/ in Iranian in general (ibid.), including in Kurmanji (cf. [hɑvin] 'summer' < PIE *sem- 'summer', [hær] 'every' < PIE *solo-, Pokorny 2007. Thus, with their continuation of Proto-Iranian *h- as [ħ], items such as [ħæft] ('seven') are exceptional. Items such as [pæħt] ~ [pæʕt] ('baked') might appear to display a pharyngeal continuation of a PIE stop ([pæħt] < *pekw-), but in fact, the PIE *k is not continued in this root: generally, Proto-Indo-Iranian *k became *x when it preceded another "non-syllabic consonant" (Fortson IV 2010: 228), only to have *x deleted in just such contexts in Kurmanji Kurdish (cf. Persian [feɾo:xt], Kurmanji Kurdish [fɪɾot], 'sold', note the lack of pharyngeal or any other consonant before [t]). The pharyngeal in [pæħt], therefore, is as lacking in explanation as the one in [tæʕv] ('sun'). Any inherited pharyngeal in Kurmanji Kurdish in an item such as [tæʕv] would imply a sound change to the effect of PIE *p > [ʕv], a sound change which can't be justified based on any data but this item. While in general the pharyngeals appear to make Kurmanji more irregular in its continuations than Persian, even in cases where the two languages are both irregular we have no clear indication of why Kurmanji has a pharyngeal: in items such as [ħɪʃk] ('dry'), Persian also has a guttural onset ([xoʃk]), which Paul (2008) terms one of the "unetymological" onsets. But this does nothing to account for the pharyngeals at large in Kurmanji. Persian also has onsets emerging *ex nihilo* as far as PIE is concerned, as in [hæʃt] ('eight') (Kurmanji [ħæʃt]), but Kurmanji also has many *ex nihilo* onset h- which are not pharyngeal, which is often the case when breaking up onset clusters, e.g. Sorani Kurdish [æsteɾæ] ('star'), Kurmanji [histeɾik] (among other forms, but all differences between the Sorani and Kurmanji forms are regular except the onset h-, which is nonetheless not pharyngeal).

It should be noted that another reason to discount the idea of inherited pharyngeals is the fact that of the items in Table 1, most shift between dialects. As previously mentioned, some dialects contain pharyngealized obstruents of the type found in Arabic, but as these do not extend to all or perhaps even most dialects, they are not included in many phonological summaries of Kurmanji. Other dialects which may share the pharyngeal sound in items like [ʕæwɾ] may disagree on whether the items 'seven' and 'eight' both begin with [ħ], or if only seven does. I propose that this shift between dialects may indicate a relatively recent process of pharyngealization of inherited vocabulary, which may even be ongoing.

Some examples of the variation can be seen in Table 2, which shows pharyngeal items alongside non-pharyngeal cognates in Avestan (the earliest attested Iranian language), Modern Persian (a closer relative), and the Mêrdînî dialect of Kurmanji, whose speakers produce many pharyngeals but strongly object to some pharyngeal forms common in other dialects. Of particular note is the correspondence between Persian [ɒː] and a vowel-pharyngeal sequence in (c) and (d). Ordinarily, Persian [ɒː] corresponds to Kurmanji [ɑ]. Kurmanji [ɑ] is here not written with contrastive length, although, like its Persian counterpart, it is generally phonetically longer. Historically, this vowel was the long form of the vowel that became /æ/, but today the salience of length for native speakers is not clear (see Haig & Öpengin 2018, who use the term "full vowels", noting the lack of phonemic length contrast). [æ], on the other hand, is not considered long in any sense, and the sequence [æʕ] appears as an innovation akin to diphthongization. In the Mêrdînî dialect in Table 2, instances of the "full vowel" [ɑ] are non-innovative. Note in particular (a), where the Persian cognate differs from the Mêrdînî dialect of Kurmanji (rather than the Persian and Mêrdînî forms being close to indistinguishable), but a more innovative pharyngeal form is still found in other dialects ([tʃæʕv]).

Existing research does not put forward any explanation for what appears to be the spontaneous emergence of pharyngeals or pharyngealization in various lexical items sometime in the historical timeline between Proto-Iranian (1500 BCE, Windfuhr 2009) and (the written attestation of) Kurdish[1]. Works that mention the pharyngeals in inherited as well as loan vocabulary fail to go beyond an ascription of their presence to contact with a Semitic language, whether Aramaic (Hoberman 1985) or Arabic (Haig & Matras 2002). The restriction of pharyngeals to those Western Iranian languages in contact with

[1]Varieties identified as Kurdish are attested from at least the 16th century (McCarus 2009).

Table 2: Selected Kurmanji pharyngeal forms with Persian and Avestan cognates

#	Avestan	Persian	Kurmanji Kurdish	Mêrdînî dialect of Kurmanji	Gloss
(a)	čašman-	[tʃʰæʃm]	[tʃæʕv]	[tʃʰɑv]	'eye'
(b)	jaθra-	[zæhr]	[ʒæħɾ] ~ [ʒæʕɾ]	[ʒæħɾ]	'poison'
(c)	mairya-	[mɒːr]	[mæʕɾ]	[mɑɾ]	'snake'
(d)	masya-	[mɒːhiː]	[mæʕsi]	[mɑsi]	'fish'

languages which natively possess pharyngeals (such as Arabic, Aramaic, and Caucasian languages) is consistent with this hypothesis. However, it does not, at least on its own, account for the pattern of realization as it actually exists, either in loan vocabulary or in inherited vocabulary in Kurmanji.

In this paper, I analyze pharyngeal sounds in Kurmanji inherited vocabulary as an example of contact-induced phonological change with a clear phonetic basis. I show that Kurmanji is relatively systematic in its phonological treatment of Arabic loan vocabulary and inherited vocabulary alike. Additionally, in contrast to the idea that there are pharyngeal consonants in the Kurmanji inventory, I argue that the distribution of these sounds is consistent with my analysis of underlying pharyngealized vowel phonemes (as rhetorically proposed, and immediately rejected, in Kahn 1976: 47), drawing on evidence from the syllable structure of the language.

I explain the historical emergence of pharyngeals in terms of phonetic recategorization of vowels and /h/ in syllabic environments involving acoustically "flat" consonants (Jakobson et al. 1952; Ohala 1985), which include sounds that are labial, pharyngeal, or retroflex (which includes to some extent rhotics and postalveolars[2]). All of these sounds share the quality of lowering the F2 of adjacent vowels. I propose that contact with Arabic facilitates the evolution of a pharyngeal category through a perceptual magnet effect (Blevins 2017). A large proportion of Kurdish speakers have historically and are still presently exposed to Arabic to such an extent as to effect widespread bilingualism. Such speakers have been hearing and producing pharyngeals for an extensive historical period. With this contact influence, Kurmanji speakers increasingly developed a category of pharyngealized vowels and/or syllables, alongside an increasing store of lexicon loaned from Arabic with

[2]Depending on the analysis, all sibilants might be classed as "flat", but in this paper only postalveolars are relevant to the analysis.

these sounds. Consequently, the conditions were created for sounds in inherited vocabulary to be reanalyzed as pharyngeal. Speech sounds which are not pharyngeal but which produce pharyngeal-like phonetic effects were and are the most likely candidates for such recategorization.

In my analysis, the result of this recategorization is that in both inherited and borrowed vocabulary, Kurmanji has developed syllables of certain restricted types which display pharyngealization. A seemingly inviolable restriction is that these syllables, regardless of their origin, are associated with one of two vowels. I posit that this restriction represents an expansion of the currently accepted vowel inventory of the language. These pharyngealized vowel phonemes might be represented as /æˤ/ and /ɪˤ/. This paper argues that all pharyngeal sounds of the language may be accounted for through reference to these two pharyngealized vowel phonemes.

In Section 2, I provide an overview of the phonological system of Kurmanji, providing justification for the aforementioned reanalysis of pharyngeals in the language. In Section 3, the evolution of pharyngeals in inherited Kurmanji words is explained in terms of the acoustic properties of pharyngeals in particular and the interaction between the perceptual magnet effect and areal sound patterns more generally (Blevins 2017), drawing from my own fieldwork and the published works of Chyet (2003) and Thackston (2006). While the majority of pharyngeals in Kurmanji occur either in Arabic loans or in inherited vocabulary as a result of the contact-induced sound change sketched above, a handful of words with pharyngeals cannot be analyzed in this way. These examples are accounted for in Section 4 as instances of non-phonetic change, motivated by analogy and sequential contamination with cardinal numbers (of the type seen across Indo-European, e.g. the initial d- of the Slavic word for 'nine' from the word for 'ten').

2 An overview of Kurmanji phonology

2.1 Vowel and consonant inventory

Kurmanji is generally analyzed as having eight vowels (Haig & Öpengin 2018; Thackston 2006). A generally accepted vowel inventory for the language is depicted in Table 3. While most of these vowels are relatively stable across Kurmanji dialects, three of them – /æ/, /ɪ/, and /ʊ/ – vary in quality across and within dialects (Thackston 2006). Of particular note is the diphthongization of /ʊ/ to [wɪ].

Table 3: Kurmanji vowel phonemes (based on Thackston 2006)

	Front	Back
High	/i/	/u/
Mid-high	/ɪ/ [ɪ] ~ [i̽]	/ʊ/ [ʊ] ~ [wɪ]
Mid	/e/	/o/
Low	/æ/ [æ] ~ [ɛ] ~ [ə]	/ɑ/

The system indicated in Table 3, with the internal variation accounted for, may also be said to incorporate the vowel system of Sorani Kurdish, one of the "points of unity" between Sorani and Kurmanji. Some examples of minimal pairs making use of these vowel contrasts can be seen in (1).

(1) Kurmanji vowel contrasts
   a) /i/ vs. /u/: /ʒiɾ/ 'clever', /ʒuɾ/ 'room'
   b) /ɪ/ vs. /i/: /dɪn/ 'other', /din/ 'mad'
   c) /ʊ/ vs. /u/: /dʒʊræ/ 'argument, disagreement', /dʒuræ/ 'type'[3]
   d) /i/ vs. /e/: /zin/ 'saddle', /zen/ 'mind'
   e) /o/ vs. /u/: /doɾ/ 'around', /duɾ/ 'far'
   f) /æ/ vs. /e/: /kʰæɾ/ 'donkey', /kʰeɾ/ 'knife'
   g) /ɑ/ vs. /æ/: /dɑɾ/ 'tree', /dæɾ/ 'outside'

An eight-vowel system of the type seen in Table 3 is expected following the shift of the Middle Iranian diphthongs to vowel phonemes (i.e. *ai > /e/ and *au > /o/). The consonantal inventory of Kurmanji, on the other hand, is significantly more complex than expected, and is not necessarily agreed upon. Part of the difficulty lies in the fact that Kurmanji speakers, often being multilingual, will note contrasts of neighbouring languages which were not historically contrastive in Kurmanji itself. For example, some sources indicate a voicing contrast for velar fricatives, as exists in Persian, Armenian, some Turkic varieties, Arabic, etc., but in Kurmanji occurs only in loan vocabulary (e.g. [ɣæzɑl], 'gazelle', which is in fact [xæzɑl for many speakers). The glottal stop, which is contrastive in Arabic and Aramaic, is sometimes counted as well, although it may only contrast in loan vocabulary from these languages in Kurmanji.

Table 4 contains those consonantal sounds for which I find ample evidence for their contrastive nature across Kurmanji in general. Pharyngeals are

[3]Chyet (2003).

omitted from this list of contrastive consonant sounds, not due to the existence of Kurmanji dialects which lack these sounds, but rather due to this paper's reanalysis of these sounds as the realization of a vocalic, rather than consonantal contrast.

Table 4: Consonant phonemes of Kurmanji

	Labial	Alveolar	Postalveolar	Palatal	Velar	Uvular	Glottal
Stop/affricate	pʰ p b	tʰ t d	tʃʰ tʃ dʒ		kʰ k g	q	
Fricative	f v	s z	ʃ ʒ		x		h
Nasal	m	n					
Lateral		l					
Flap		ɾ					
Trill		r					
Glide	w			j			

The three-way contrast between voiceless, aspirated, and voiced stops (including the affricate series) seen in Table 4 is a feature of many dialects of Kurmanji. An example of a minimal triplet may be found in /phiɾ/ 'religious elder', /piɾ/ 'old woman', and /biɾ/ 'memory'. The aspiration contrast is only found in simple onsets; elsewhere, the only laryngeal contrast is that of voicing.

As is the case across a broad geography of Iranian languages[4], there is also a uvular stop which does not contrast for laryngeal features[5], which is contrasted with the other three dorsal stops, as seen in (2).

[4] In addition to several Western Iranian languages, /q/ is also found, without laryngeal contrast, in Yaghnobi (Bird 2007; Khromov 1972), Ossetian (in Iron dialect it regularly continues Proto-Iranian *g-, Thordarson 1989: 464), and most Pamiri languages (Edelman & Dodykhudoeva 2009: 779–780).

[5] This sound appears predominantly in Arabic and Turkic loan vocabulary. However, it is also to be found in some inherited vocabulary, e.g. [phɑqɪʒ], 'clean', cognate to Persian /pɒ:ki:zæ/, 'tidy', and would be a worthy subject of study on its own.

(2) Uvular stop contrasted with the velar stops
   a) /q/ vs. /k/ /qæɾ/ 'debt', /kæɾ/ 'piece'
   b) /q/ vs. /kʰ/: /qædæɾ/ 'fate', /kʰædæɾ/ 'worry'
   c) /q/ vs. /g/: /qændʒ/ 'good', /gændʒ/ 'young'

The analysis in Table 4 differs from other analyses of Kurmanji phonemes (such as Haig & Öpengin 2018; Thackston 2006) in not including the velar nasal [ŋ] as a phoneme. This is because [ŋ] only occurs syllable-finally, where it alternates with [ŋg] whenever followed by a vowel. I therefore analyze it as a coda realization of a cluster /ng/ (see Section 2.2).

Note that Kurmanji lacks contrastive geminates, or any true contrastive length for vowels or consonants. This is particularly noteworthy given that consonantal gemination is a feature of the phonology of several languages with which Kurdish is in long-standing contact, including Persian and Arabic.

2.2 Syllable structure and phonotactics

Relevant to my reanalysis of Kurmanji pharyngeals as emerging from an underlying feature is the structure of the syllable, in particular with regard to vowels, clusters and glides: the distribution of pharyngeals in terms of adjacent vowels figures heavily into my analysis, as does their relationship to clusters compared to consonants on the one hand and glides on the other.

Every Kurmanji syllable necessarily consists of one vowel (no other sonorant may serve as the nucleus). Syllables without phonological onsets do occur, such as those examples in Table 5. These are often low vowels, with /e/, /i/, and /o/-initial words being less common, /ɪ/ being rarer still, and no words unambiguously beginning with /ʊ/ or /u/ (see Table 5, as well as Chyet 2003: 282, 631–632 especially noting the other possible forms of these words). While most vowels may appear possible word-finally, /ɪ/ and /ʊ/, the mid-high vowels, are effectively limited to clitics, and /o/ is limited to a single suffix in most dialects (see Table 6). CVC syllables are common with all vowels, as shown in Table 7.

Table 5: Syllable-initial vowels in Kurmanji

#	Syllable type	Example word	Gloss
(a)	VC	[ɑv]	'water'
(b)	VC	[æv]	'this'
(c)	VC	[eʲ]	'pain'
(d)	VC	[ɪsteɾɪk]	'star'
(e)	VCC	[isk]	'hiccup'
(f)	VC	[ol]	'religion'
(g)	VCC	[ʊlm]	'science'
(h)	V	[u]	'and'

Notes d): Epenthetic initial vowel, more conservative form is [steɾɪk], further innovation with h-initial [hɪsteɾɪk] is also common.
Notes g): More conservatively, this Arabic loan may be [ʕɪlm], which is not phonetically vowel-initial. Words generally do not begin with [ʊ].
Notes h): Enclitic, words generally do not begin with [u].

Table 6: Syllable-final vowels in Kurmanji

#	Syllable type	Example word	Dialectal variants	Gloss
(a)	CV	[bɑ]		'wind'
(b)	CV	[dʒæ]	[dʒæh] ~ [dʒæhæ]	'barley'
(c)	CV	[te]	[dɪhe]	'comes, is coming'
(d)	CV	[tɻɪ]		'what'
(e)	CV	[si]	[sih]	'shadow'
(f)	CV	[zu]		'quick, early'

Notes b): The CV form is innovative, but common. A copy vowel and consequent new syllable can result from preservation of the -h.
Notes c): [te] is the more innovative form, but it is very widespread.
Notes d): Most of the few words ending in [ɪ] are enclitics, multisyllabic words cannot end in [ɪ], this word is therefore exceptional.
Notes e): The h-coda is the more conservative form. I am unaware of a copy-vowel form, cf. (b).

Table 7: CVC monosyllabic words in Kurmanji

#	Syllable type	Example word	Gloss
(a)	CVC	[bɑɾ]	'load'
(b)	CVC	[mæt]	'paternal aunt'
(c)	CVC	[zer]	'gold'
(d)	CVC	[gɪl]	'mud'
(e)	CVC	[ʃin]	'blue'
(f)	CVC	[dʒʊx]	'yoke'
(g)	CVC	[qum]	'sand'

All consonants in Table 4 except the flap /ɾ/ may appear as simple onsets, although /j/ is rare word-initially. In the analysis of Kurmanji as possessing two pharyngeals (/ʕ/ and /ħ/), these too are permissible onsets. All consonants except the aspirated obstruents in Table 4 may appear as simple codas. The pharyngeals do not generally appear as simple codas word-finally, and the general context for post-vocalic pharyngeals is intervocalic (which may be analyzed as onset).

In addition, syllables may end in consonant clusters, and, according to Haig & Öpengin (2018: 170), begin with them. Whether a given cluster is permissible or not is inconsistent across Kurmanji dialects, with some allowing for almost no clusters, breaking them up via regular processes of epenthesis. As this appears to have little bearing on pharyngeals in the language, this is not treated in detail here.

In general, the syllable structure of Kurmanji is under-analyzed (Öpengin, personal correspondence). Karimi-Doustan (2002) gives the maximal syllable for Kurdish as (C)(C)V(C)(C). To this I would add that there are two items ([stɾɑn] 'song' and [stɾu] 'horn') which apparently allow for a [stɾ] onset cluster, bringing the theoretical maximal syllable to (s)(C)(C)V(C)(C). Examples of Kurmanji syllables containing various clusters are shown in Table 8 and Table 9.

As previously mentioned (in Section 2.2), simple onsets may consist of any consonant except the flap /ɾ/. In onset clusters consisting of two consonants, these may either be obstruent-liquid clusters, such as [dɾeʒ], or clusters which consist of a fricative followed by a non-fricative consonant, such as [spi] or [ʒmɑɾ]. The maximal three-consonant onset seems to invariably consist of [stɾ].

Table 8: Final clusters involving stops, both oral and nasal

#	Syllable type	Example word	Gloss
(a)	VCC	[æwk]	'thingamajig'
(b)	VCC	[ɑrd]	'flour'
(c)	VCC	[isk]	'hiccup'
(d)	CVCC	[ʂæjb]	'shame'
(e)	CVCC	[dʒæʒn]	'festival'
(f)	CVCC	[k^{h}æsk]	'green'
(g)	CVCC	[kævn]	'old' (of things)
(h)	CVCC	[ɲærm]	'shame'

Table 9: Initial clusters involving stops, both oral and nasal

#	Syllable type	Example word	Gloss
(a)	CCV	[spi]	'white'
(b)	CCVC	[ʒmɑr]	'number'
(c)	CCVC	[dɾeʒ]	'long'
(d)	CCVCC	[bɾusk]	'lightning'
(e)	CCCV	[stɾu]	'horn' (of an animal)
(f)	CCCVC	[stɾɑn]	'song'

Simple codas may consist of any consonant (although in many dialects /h/ is an exception). Coda clusters consist of any non-stop consonant (including glides) followed by any fricative, or (non-aspirated) oral or nasal stop.

In both onset and coda clusters, there are no known instances of clusters containing an aspirated stop, or containing two fricatives. Additionally, affricates may not occur in initial clusters. Obstruent-obstruent clusters are voiceless, while obstruent-nasal clusters are voiced (e.g. [dʒæʒn], 'festival', c.f. Persian /dʒæʃn/).

The Sonority Sequencing Principle states that sonority must not rise between the syllable peak and any other part of the syllable (Blevins 1995). This principle is violated in two ways in Kurdish. Firstly, in common with other Iranian languages, nasal stops can follow voiced fricatives in the coda (e.g. [dʒæʒn], [kævn] in Table 8). Secondly, in common with other Indo-European languages, /s/ may precede lower sonority sounds in onset position (e.g. [spi], [stɾɑn] in Table 9), a trend which seems to have been extended to other fricatives in Kurdish, particularly as one moves southwards. Indeed,

onset cluster constraints become more relaxed the further south one goes, to the point where initial stop-stop clusters are claimed to be accepted in at least some varieties of Sorani Kurdish, e.g. [kteb] (Haig & Öpengin 2018: 170).

2.3 Glides and pharyngeals in Kurmanji

Some analyses of Kurmanji refer to the pharyngeals as "fricatives" (Haig & Öpengin 2018; Thackston 2006). If this is taken at face value, it raises a third issue for the sonority sequencing principle, as pharyngeals may precede sonorants in coda position, as in [bæħɾ], also pronounced [bæʕɾ] (see Table 10, where all post-vocalic instances of [ħ] may be voiced). This would mean that, in addition to the permissibility of coda fricative-nasal clusters, pharyngeal "fricatives" on their own would allow for another violation of the Sonority Sequencing Principle, with sonorants in general. If pharyngeals were indeed fricatives, they would pattern unusually in another respect, in that they would be the only fricatives capable of forming fricative-fricative clusters (see [tæʕv] in Table 1, [tʃæʕv] in Table 2). This may, however, simply be a misnomer of convenience. Esling (2010: 695) explains the lack of contrast between pharyngeal approximates and fricatives cross-linguistically through a lack of acoustic salience of such a contrast, combined with the articulatory difficulty in producing pharyngeal frication. With such a distinct class of pharyngeal fricatives being hypothetical, it is most likely that, in keeping with the sonority constraints of Kurdish, pharyngeals are indeed glides or approximants, and not fricatives.

If the pharyngeals are indeed glides and not fricatives, their purported voicing contrast would make them the only sonorants in the language with such a contrast. However, a real voicing contrast for pharyngeals in Kurmanji is not obvious. Firstly, there is the lack of any minimal pairs, either in onset position where the "contrast" is usually noted, or in coda clusters where the posited voiceless /ħ/ is often realized as voiced [ʕ], even when the following consonant is a voiceless obstruent, as in [pæʕt] ~ [pæħt]. This post-vocalic pattern mirrors the sonorant /h/, which is not held to contrast for voicing in any context, but phonetically has variable voicing post-vocalically. Indeed, post-vocalically the voiced form of the pharyngeal is more common, with the voiceless form almost invariably produced with an epenthetic vowel of some sort, in common with the treatment of etymological coda /h/ more broadly (recall [dʒæh] ~ [dʒæhæ] in Table 6). In the analysis which follows, I suggest that [ħ] is the surface realization of an underlying /h/ phoneme, produced in the environment of a *pharyngealized vowel*.

Table 10: Coda clusters containing glides and pharyngeals

#	Syllable	Example word	Gloss
(a)	CVCC	[ɤæwɾ]	'cloud'
(b)	CVCC	[ræwʃ]	'condition'
(c)	CVCC	[p^{h}æjv]	'word'
(d)	CVCC	[ħæwl]	'effort'
(e)	CVCC	[ħæjf]	'revenge'
(f)	CVCC	[bæʕɾ] ~ [bæħɾ]	'sea'
(g)	CVCC	[ʒæʕɾ] ~ [ʒæħɾ]	'poison'
(h)	CVCC	[pæʕt] ~ [pæht]	'baked' (past stem)

2.4 Patterning of Pharyngeals in Kurmanji

In order to ground a reanalysis of phonetic pharyngeals as the realizations of pharyngealized vowel phonemes, it is crucial to investigate the distribution of pharyngeals relative to the vowels. Most pharyngeals in inherited vocabulary are directly adjacent to the vowel /æ/, with a few adjacent to /ɪ/. No such vocalic constraint applies to any of the other consonants or glides in the language.

Most of the inherited [æ] syllables with a pharyngeal contain a labial (see Table 1 and Table 2), and the [ɪ] syllables contain postalveolars. The word for 'paradise', [bæħɪʃt] (Table 2) contains one of each syllable type on either side of the pharyngeal. Evidence for this from the vocabulary and explanations for exceptional cases will be provided in Section 3.

Crucially for the plausibility of this position, in its phonetic realization, a pharyngealized vowel must generate a pharyngeal in a permissible syllabic position for a glide. All phonetic instances of [ħ] are analyzed as the realisation of an underlying pharyngealized vowel phoneme in the environment of an underlying /h/ phoneme, which together generate the surface [ħ]. With the [ħ] sound, this takes the place of an underlying /h/. Note that like /h/, /j/, and /w/, the pharyngeals do not seem to occur in onset clusters (see Table 10).

3 Pharyngeals in Kurmanji

3.1 Motivation for pharyngeals in inherited vocabulary

Kurmanji presents a puzzle with two inter-related parts. Firstly, how did inherited vocabulary items such as those in Table 1 and Table 2 come to contain pharyngeals? More specifically, how did this happen when no specific sound change from an earlier Iranian language can be proposed to account for the pharyngeals? Secondly, why are pharyngeals restricted in their distribution with regard to adjacent vowels, as described in Section 2.4?

Table 11: Some hypothetical but unattested pharyngeals in inherited vocabulary

#	Actual item	Unattested pharyngeal form(s)	Gloss
(a)	[ɑgɪɾ]	*[ʕɑgɪɾ], *[ʕægɪɾ]	'fire'
(b)	[dʊr]	*[dʊʕr], *[dɪʕr]	'pearl'
(c)	[gædæ]	*[gæʕdæ]	'vagrant'
(d)	[hek]	*[ħek]	'egg'
(e)	[hostɑ]	*[ħostɑ]	'expert'

In order to answer these questions, it is useful to consider the absence of pharyngeals in other contexts. Note that all of the items in Table 11 could contain pharyngeals and broadly conform to Kurmanji syllable structure, but the pharyngeals should surface in the context of an appropriate vowel adjacent to a labial consonant. As mentioned in Section 2.4, pharyngeals must be adjacent to the vowels [æ] or [ɪ], while in the hypothetical items *[ʕɑgɪɾ], *[dʊʕr], *[ħek] and *[ħostɑ], they are not. The lack of attestation of innovative forms such as *[ʕægɪɾ] (for 'fire') contrasts with attested innovative forms such as [mæʕɾ], which exist alongside forms such as [mɑɾ] (see Table 2 in Section 1).

Furthermore, the pharyngeal forms in Table 11, including those with an appropriate vowel, such as *[gæʕdæ], may be ill-formed because they lack a labial consonant adjacent to the pharyngeal. This paper seeks to explain not only the presence of pharyngeals in inherited vocabulary items where they do occur, but also their absence in items such as those in Table 11, where they do not.

3.2 Arabic origins of the vocalic constraint

The first constraint on pharyngeals in inherited vocabulary is, as previously mentioned, that pharyngeals must be adjacent to one of two vowels, [æ] or [ɪ]. This constraint seems to be inviolable, even if other vowels are observed in more conservative, non-pharyngeal forms. Some examples of such dialectal innovations, of a pharyngeal emerging with a shift in the vowel to accommodate it, may be seen in Table 12, where pharyngeal forms replace the conservative forms vowel with the permissible form. [ɑ] shifts to [æ], and [ɪ] replaces a round vowel in the non-pharyngeal Sorani forms. The [o] in [hoʃ] is transparently a different vowel, for the treatment of [wʊ] ~ [wɪ] as a vowel unit, recall that these forms are analyzed as allophones of /ʊ/. [wʊ], [wɪ], and [ʊ] do not contrast in Kurmanji or Sorani Kurdish, as all apparently represent a single round or labialized mid-high vowel phoneme. Word-initially, [wɪ] is common in transcriptions of Kurmanji, while [wʊ] is common in phonetic transcriptions of Sorani. In both varieties, [ʊ] is the allophone which surfaces in non-initial position (see Table 3).

Table 12: Vowel shift in pharyngeal vs. non-pharyngeal contexts

#	Pharyngeal form	Non-pharyngeal form	Context notes	Gloss
(a)	[bæʕdʒɑn]	[bɑdʒɑn]	Shift occurs within Kurmanji	‘eggplant’
(b)	[bæʕlif]	[bɑlif]	Shift occurs within Kurmanji	‘pillow’
(c)	[ħɪʃ]	[hoʃ]	Non-pharyngeal form is Sorani	‘wits, reason’
(d)	[ħɪʃk]	[wɪʃk] ~ [wʊʃk]	Non-pharyngeal form is Sorani	‘dry, arid, solid’
(e)	[mæħin]	[mɑhin]	Shift occurs within Kurmanji	‘mare’
(f)	[tʃæʕv]	[tʃʰɑv]	Shift occurs within Kurmanji	‘eye’

The vowel pairs in question fall into natural classes; the two low vowels, /æ/ and /ɑ/, and the two mid-high vowels, /ɪ/ and /ʊ/ (see Table 3), are paired for merger into the two pharyngealized vowels which share these features. In pharyngeal syllables, the low vowels merge into a single pharyngealized low vowel, which I write /æˤ/, although the choice of “æ” is arbi-

trary and not based on a particular "frontness": it could just as easily be analyzed as /aˤ/ or /ɐˤ/[6], but the orthographic low vowel in pharyngeal syllables in written Kurmanji tends to be the same symbol used for /æ/, implying an indigenous association (hence the [æ]s throughout this paper). Likewise, the two mid-high vowels merge into a single pharyngealized mid-high vowel /ɪˤ/.

But why the association between the vowels /æ/ and /ɪ/ and pharyngealized syllables in the first place? Arabic, the language from which the pharyngeals are held to have spread, has a three-vowel system with a two-way length contrast (/a/, /i/, and /u/, short and long). The exact quality of Arabic vowels varies between Arabic dialects, and (as happened historically in Iranian languages as well) differences in vowel length have effected changes in vowel quality.

The short vowels in Arabic loanwords are relatively straightforward. In Kurmanji, Arabic /a/ is realized as Kurmanji /æ/, and, in my analysis, in pharyngeal syllables it is realized as the pharyngealized low vowel phoneme /æˤ/. At first glance, it would appear that Arabic short /i/ and /u/ merge into Kurmanji /ɪ/. But given that the Arabic dialects with which Kurmanji is in contact are generally North Mesopotamian, "the majority" of which merge earlier /i/ and /u/ into a single schwa phoneme (Watson 2002: 21, citing Jastrow 1980: 54), it might be more accurate to say that Kurmanji has taken this dialectal Arabic schwa phoneme in as /ɪ/. The "two-short-vowel system" of Arabic dialects (Watson 2002: 22) perfectly mirrors the two-pharyngeal-vowel system I propose here. The greater frequency of pharyngeal /æˤ/ than pharyngeal /ɪˤ/ may partially be explained by the fact that /ɪ/ is the default epenthetic vowel in Kurmanji; and/or by its frequency in Arabic loans being augmented by modern Arabic final shortening (Holes 2004: 61) in items such as [inʃælæ] ('God willing') < Qur'anic Arabic /inʃaːʔałłaːh/, [mæˤnæ] ('meaning') < Qur'anic Arabic /maˤnaː/.

The Arabic long vowels present their own puzzle. As in Arabic, whose long /iː/ is realized as [i] in Kurmanji, there is no contrast between /ɪj/ (Arabic /ij/) and the "full vowel" /i/. This is in contrast to the uniformity of the triangular vowel system of Arabic as it is usually presented, and indeed presents a similar issue of uniformity of Kurmanji vowels, unless we can demonstrate that Kurmanji "full vowels" are all perceptually indistinguishable from diph-

[6] Likewise there is no significance to the pharyngeal symbol being positioned after the vowel in the underlying form. It could just as easily be before or on top of the vowel, as the vowel itself has a pharyngeal association.

thongs[7]. But under this assumption, how would the other Arabic long vowels be analyzed? Can Arabic long /u:/ be phonemically /ʊw/ to Kurmanji speakers? Combined with the [wɪ] allophone of /ʊ/, this would imply that Kurmanji speakers cannot perceive a contrast between initial [u] and [wɪw], or final [u] and [ʊw][8]. Finally of course, there is the regular correspondence of Arabic /a:/ to Kurmanji /ɑ/, which contrasts with both /æ/ and /æˤ/ in Kurmanji.

Even leaving aside the anomalous nature of Kurmanji /i/ (and Arabic /i:/), Arabic long vowels still present a unique problem for pharyngeals. So far as I am aware, the inherited vocabulary of Kurmanji lacks a single example of a pharyngeal whose only adjacent vowel is a "full vowel". Of the Kurmanji "full vowels", two (/e/ and /o/) can be assumed to lack an association with pharyngeals because these phonemes are not to be found in many varieties of Arabic, and are certainly absent in Qur'anic Arabic. The remaining three "full vowels" (/ɑ/, /i/, /u/), however, are all known to regularly correspond to the Arabic long vowels (/a:/, /i:/, /u:/). As the Arabic language is rich in pharyngeals and possesses a templatic morphology which does not allow for a different space in the syllable structure for pharyngeals than for, e.g. oral stops, this would not appear to suggest any problem for a lack of pharyngeal association with these vowels. So what has become of Arabic loans with a pharyngeal adjacent to an Arabic long vowel and no short vowel?

A significant portion of the vocabulary is eliminated by the requirement to not have a short vowel on either side of the pharyngeal, but several common monosyllabic items may still be offered as evidence. From front to back and top to bottom, I will present an example for each of the Arabic long vowels in terms of the colloquial Arabic and the Kurmanji realizations.

For Arabic /i:/, an unavoidably common item is /ʕi:d/, 'holiday'. Loaned into Kurmanji, an epenthetic vowel [æ] is not only observed phonetically but standardised in orthography, implying a salience to the [æ] vowel, while the historical nucleus /i/ has become the coda glide /j/: [ʕæjd] (note Chyet 2003: 184, 283: no <'îd> form is found in the <E> or <Î> sections). Vowel lowering in a pharyngeal context is a pattern in Arabic (Watson 2002: 46), and a diphthongisation of /i:/ similar to the Kurmanji pattern is attested even in South Semitic languages, e.g. Mehri, a South Arabian language (Watson

[7] In fact, we do not find such a pattern. Other than /i/, the other "diphthong equivalent" vowel is the round mid-high vowel /ʊ/, whose unround equivalent has no diphthong allophone, just as the high round vowel /u/ does not.

[8] However absurd or plausible this may sound to the reader, it is an empirical question that may be tested empirically.

2012: 26), such that we may claim the reason for a lack of association between Kurmanji /i/ and pharyngeals is that this pattern had already become dominant in the dialects of Arabic with which Kurmanji was in contact. Potential counter-examples could represent dialects in closer contact with Arabic dialects which allow for such sequences, or could simply represent etymological spellings or learned pronunciation. But in any event, this pattern is expected in articulatory terms, as /i:/ is a front and high vowel, in contrast to the low and back quality of pharyngeal sounds, such that the mouth must "travel through" an intermediary to reach its articulatory target. Thus, Kurmanji speakers by and large would only hear and only produce a pharyngeal with an intermediary vowel in such contexts.

The other two Arabic long vowels, /u:/ and /a:/, are both back, and /a:/ is low. In these cases, we should not expect such a strong need for the pharyngeal to "travel". Some distinct Kurmanji treatment of these vowels is widespread, for example Arabic /ru:ħ/ > Kurmanji [rıħ]. However, this is not the general trend. Note for example, the Arabic /ħu:t/ ('whale'), whose long vowel is preserved in the expected "full vowel" form with the pharyngeal lost: [hut][9]. Likewise, the extremely common Arabic /ħa:l/ ('status') is reported without a pharyngeal, but much more frequently it is reported with pharyngeal and the "full vowel". My own doubts about this pronunciation aside (I am convinced I only ever hear [hɑl] or [ħæl]), this is one of the few Arabic loan items on which consultants of various dialect backgrounds agreed on its pharyngeal onset and "full vowel" nucleus.

Variation exists across the language with regard to the pharyngeals as with other features. But in both inherited vocabulary and Arabic loan vocabulary, pharyngeal syllables are overwhelmingly [æ]-nucleic, and to a lesser extent [ɪ]-nucleic. This distribution motivates my analysis of two additional vowel phonemes, namely the pharyngealized vowels /æˤ/ and /ɪˤ/. If pharyngeal-[ɑ] syllables are indeed present in items such as [ħɑl], a third /ɑˤ/ phoneme would need to be posited.

It is noteworthy that in dialects such as that of Mêrdîn (see Table 2), which are adjacent to large Arabic-speaking populations to this day, these pharyngeals (or pharyngealized vowels in this analysis) have penetrated the inherited vocabulary to a lesser extent than in dialects such as that of Qers, which sits on the outskirts of the Kurmanji-speaking area, and where the neighbor-

[9]Except in two Soviet sources, in opposition to all non-Soviet sources consulted by Chyet 2003: 266, 281. Counter-intuitive though this would be, it would imply a dialectal outlier in Caucasian Kurdish, which may have been in contact with the related Tat language, with a similar vocabulary but distinct rules for pharyngeals (including in inherited vocabulary).

ing languages are pharyngeal-free varieties of Turkish and Armenian. Possibly, in the absence of reinforcement from Arabic speakers, the distinction between inherited and Arabic loan vocabulary has become more blurred, as bilingualism shifted from Kurdish/Arabic to Kurdish/Turkish. The more that pharyngeal sounds are associated with the native phonology rather than a stratum of vocabulary, the more the Arabic pharyngeal consonants fade away, and the more that speakers mainly hear the pharyngeal effects on the vowel, rather than the brief "consonantal" segment.

3.3 Phonetic motivation for pharyngeal syllables: "Flat" consonants and formants

In addition to my proposed constraint on adjacent vowels, another apparent constraint on pharyngeals in inherited vocabulary is that the pharyngealized vowel must be adjacent to a "flat" consonant. "Flat" consonants include both labials and pharyngeals (Jakobson et al. 1952; Ohala 1985), which share the quality of lowering the F2 of the adjacent vowel. This phonetic effect has been observed to effect categorical changes in adjacent vowels. For example, in Chilcotin, flat consonants result in allophonic tongue root retraction on adjacent vowels (Cook 1993). This follows, since retraction of the tongue root is a feature of pharyngeal articulation (Esling 1999) and pharyngeal muscles must be contracted to produce retracted vowels and consonants (Fulop et al. 1998).

It is to be noted that the majority of pharyngeals in the inherited vocabulary of Kurmanji Kurdish are found in the coda of [æ]-syllables with a labial onset, or in the onset of an /æ/-syllable with a labial coda (see items in Table 1, Table 2 and Table 10). This shows not only a strong association between labials and pharyngeals, but also that the association must cross the syllable (through its nucleus, the vowel). Further, pharyngeals have the quality of raising F1 (Ghazeli 1981), which means that the "flat" effect might be more perceptually salient adjacent to lower vowels, like /æ/, which possess a higher F1.

The labial-pharyngeal association is also not unique to Kurmanji. In "a number of modern Arabic dialects", it has been observed that "labialization" in the form of "lip-protrusion or lip-rounding" is an "enhancing feature" for pharyngeals and pharyngealized oral consonants (Watson 2002: 269). Speakers of genetically and geographically diverse languages which lack pharyngeal articulation natively have been observed to substitute labialization for pharyngealization in Arabic words (Jakobson et al. 1952: 31, Holes 1995: 56).

Cross-linguistically, very few languages contrast labialization and pharyngealization, presumably due to their perceptual similarity (Blevins 2004: 136).

In the following section, I suggest that this perceptual similarity has facilitated a pharyngeal articulation in most of those items which today contain a pharyngealized vowel in inherited Kurmanji vocabulary.

3.4 The perceptual magnet effect

The rarity of the areal diffusion of a feature such as "pharyngeal" notwithstanding, the process by which such a phonological feature might spread is quite common. This is the perceptual magnet effect, which is frequently involved in patterns of areal sound change (Blevins 2017: 98). Blevins's hypothesis is as follows:

> Areal sound patterns are due to perceptual magnet effects within one language, where the perceptual magnets themselves are sounds from another language. As a consequence, their evolution may mimic that of internal phonetically based sound change.

In the case of Kurmanji Kurdish, the sounds from another language are the pharyngeals of Arabic, which have entered into Kurmanji Kurdish through widespread bilingualism or language shift with Arabic, a phenomenon currently absent from most other Iranian languages, in spite of heavy lexical borrowing from Arabic in earlier historical periods. As a consequence of the perceptual magnet effect, Kurmanji Kurdish speakers articulated most pharyngeal Arabic borrowings with some form of pharyngeal articulation, interpreted as pharyngeal syllables centered on certain vowels (identified and explained in Section 3.2). With these syllables present in the language, they could then act as perceptual magnets for similar syllables in the inherited vocabulary. Over time, Kurmanji Kurdish speakers identified certain syllables as pharyngeal based on phonetic criteria, such as the perceptual similarity of the formant frequencies of labial-adjacent low vowels with pharyngeal-adjacent vowels. This hypothesis explains not only the pharyngeal syllables identified with low vowels and labials, but also the few cases of inherited items with a high vowel and no labial, which also have an apparent phonetic motivation with another flat consonant type, the postalveolars.

3.5 [ɪ] and postalveolars

Most pharyngeals in both inherited and Arabic loan vocabulary are found in syllables with an [æ]-nucleus. The minority of other cases are [ɪ]-nucleic, which in inherited vocabulary are all followed by an postalveolar in the syllable coda, in addition to the [ɪ] corresponding to a round vowel in other Iranian varieties. These items might gain a pharyngeal association due to a conspiracy of phonetic factors, including:

1. an originally rounded vowel, /ʊ/, where lip-rounding represents a possible perceptual feature confusable with pharyngealization;
2. the articulatory phonetics associated with tongue root retraction, which may be present in mid-high vowels, particularly pharyngeal constriction (Fulop et al. 1998);
3. the F2-lowering effect of /ʃ/, which, while less pronounced than that of labial stops, may be significant in these contexts.

None of these three properties appears sufficient to lead to reinterpretation as pharyngealization by Kurmanji speakers on its own: one does not, for example, see mid-high vowels developing pharyngealization across inherited vocabulary, nor do most instances of /ʃ/ or /tʃ/ result in pharyngealization. However, when these features are found together in one syllable, as in the items in Table 13, they seem to have such an effect.

Table 13: [ħɪ]-initial syllables in the inherited vocabulary of Kurmanji

Kurmanji Kurdish	Sorani Kurdish	Persian	Gloss
[ħɪʃ]	[hoʃ]	[hoːʃ]	‘intellect’
[ħɪʃk]	[wɪʃk] ~ [wʊʃk]	[xoʃk]	‘dry, arid, solid’

4 Exceptional cases

4.1 Arabic words

As the source of the pharyngeals in Kurdish in general is held to be contact with Arabic, it comes as no surprise that most pharyngeals in Kurmanji Kurdish are to be found in Arabic words, and conversely, that Arabic pharyn-

geals are preserved when they are loaned into Kurmanji. This generalization, however, fails to capture the full extent of the facts. Table 14 illustrates several cases of Arabic-origin items in Kurmanji which either contain a pharyngeal that was not in the Arabic (such as Arabic /maʔmuːɾ/ > Kurmanji [mæʕmuɾ] 'officer, official'), do not contain a pharyngeal that was present in Arabic (such as Kurmanji [hæɾem] 'region', ultimately from Arabic /ħaɾiːm/), or exhibit a pharyngeal in a different part of the syllable than in the Arabic (such as Kurmanji [ʕærd] 'ground' < Arabic /ʔardˤ/).

Table 14: Arabic items in Kurmanji which contain a pharyngeal in one or both languages

#	Kurmanji	Alternate forms	Arabic	Pharyngeal change?	Gloss
(a)	[ʕærd]	[ʕærz]	/ʔardˤ/	✓	'ground'
(b)	[ʕɪlm]	[ʊlm], [ʕælm]	/ʕilm/		'knowledge'
(c)	[hæɾem]		/ħaɾiːm/	✓	'region' (Arabic: 'harem')
(d)	[ħæq]		/ħaqq/		'right(s), truth'
(e)	[mæʕmuɾ]	[mɑmuɾ]	/maʔmuːɾ/	✓	'official, officer'
(f)	[mæʕnæ]	[mɑnæ]	/maʕnaː/		'meaning'

Most of these forms are equally consistent with Arabic items being reanalyzed in accordance with a Kurmanji syllable structure which recognizes pharyngealized vowels or syllables, as they are with pharyngeal consonants in Kurmanji. Retention of pharyngeals in position, as in [ʕɪlm] ~ [ʕælm], [ħæq], or [mæʕnæ] in Table 14, could equally result from the analysis of pharyngeal consonants or from my reanalysis of pharyngealized vowels. For example, the expected consonant-for-consonant loan form of Arabic /ħaqq/ in a Kurmanji with pharyngeal consonants is [ħæq]. An underlying /æˤ/ as the nucleus of a /h/-onset, /q/-coda syllable would result in /h/ being realized with its [ħ] allophone in my analysis in Section 2.4. Likewise, if the underlying form for 'knowledge' is /Vˤlm/, the only syllabic slot for the pharyngeal is the onset, its position in the original Arabic. The same may be said for the coda position of a pharyngeal predicted for an underlying /mæˤnæ/ ('meaning').

[ʕærd] ('ground') is a different case, however. The change is quite extreme from the original Arabic, with a pharyngeal as the onset approximant replac-

ing of the pharyngealization of the coda obstruent. This coda, while permissible in Arabic phonology, is not in line with the syllable structure of most Kurmanji dialects, which lack pharyngealized obstruents as phonemes and cannot have approximant-final clusters, or any three-consonant clusters. The [ʕærd] form, however, is the expected result of the Arabic pharyngealized obstruent's F2 lowering effect on the adjacent vowel being interpreted as a pharyngealized vowel phoneme (/æˤɾd/), which would generate an onset [ʕ] in the surface form.

At first glance, [mæʕmuɾ] ('officer, official') appears to contain a pharyngeal approximant which has replaced an Arabic glottal stop. As Kurmanji is not generally held to have a contrastive glottal stop, likely this syllable was reanalyzed as pharyngeal due to the presence of /m/, a labial, both before and after the vowel, in line with the strong perceptual similarity of the flattening effect of labials and pharyngeals on adjacent vowels. One of the few other Arabic items in Kurmanji with a labial on each side of a vowel is the common men's name *Muhammad*, sometimes pronounced [mæʕmæd]. In my analysis, both of these items may be analyzed as beginning with /mæˤm/ in Kurmanji, with a single pharyngealized vowel phoneme /æˤ/ between the two nasals.

[hærem] ('region') is predicted as a pharyngeal owing to the presence of a permissible vowel next to an original pharyngeal sound in Arabic. Arabic /ħar/ syllables ought to be realized as [ħæɾ] in Kurmanji. Despite the ultimately Arabic source, this item may be a loan through Persian. The Arabic source word /ħaɾiːm/ translates more or less to 'harem' in English: compare the Persian /hæɾiːm/ may be used with a meaning closer to 'sanctum' or 'frontage'. Indeed, with the non-'region' meanings, this Arabic root's Kurmanji descendants do surface with pharyngeal articulation (e.g. Arabic /ħaraːm/ > Kurmanji [ħærɑm], 'forbidden').

4.2 Exceptional cases – analogy and contamination

A small number of inherited words in Kurmanji contain pharyngealized sounds for reasons not explainable by the phonetic principles outlined in Section 3. We must account for these items through an alternative account.

In the item [ʒæħɾ] ~ [ʒæʕɾ] ('poison', Table 2 and Table 9), no labial is to be found that might explain the pharyngeal quality of this item, which is widespread among Kurmanji dialects. A particularly rare syllable type may have strengthened the perceptual magnet effect on the few items with this common syllable structure. Three Kurmanji items of which I am aware end

in -VhR (all of them with the same vowel, [æ]), two of which are inherited ('poison' and [tæħl] ~ [tæʕl], 'bitter'), and the third is a common Arabic loan item: [bæħr] ~ [bæʕr] ('sea', also used for lakes in some varieties). The two native items with a similar structure (/tæhl/ and /ʒæhr/) may have undergone analogical change as a result, so that these three lonely friends became more similar: -æˤhR.

Another item without a labial but with a widespread pharyngeal pronunciation is [ħæʃt] ('eight', Table 1). [ħæʃt] was likely pharyngealized due to sequential contamination[10] by [ħæft], 'seven'. The latter is hardly a bold claim, as the initial /h/ in the West Iranian item for 'eight' is itself originally a case of contamination from the item for 'seven' in the first place (note the lack of any initial in the Proto-Indo-European *oḱ-tō(u), see Table 1). Under this account, the relationship between the initials in 'seven' and 'eight' from Old Iranian (e.g. Avestan) to New West Iranian broadly mirrors that between Sorani and Kurmanji Kurdish (see Table 15).

Table 15: 'Seven' and 'eight' in various Iranian varieties

Avestan	Persian	Sorani Kurdish	Kurmanji Kurdish	Gloss
hapta	[hæft]	[ħæwt]	[ħæft]	'seven'
ašta	[hæʃt]	[hæʃt]	[ħæʃt]	'eight'

Contamination of numerals is also attested in other numeral sequences in Kurmanji, with 'twelve' and 'sixteen' both containing unetymological nasalization of a vowel, in both cases following numerals in sequence with nasalized vowels that are the result of a post-vocalic /n/ at an earlier stage of the language, as illustrated in Table 16. While this implies the existence of a nasalized vowel phoneme in the language, this is quite marginal, attested only in these numerals.

Note that just as with the glottal or pharyngeal in the word for 'eight' in Table 15, a phonological feature (nasalization) is carried forward to the subsequent numeral in the count.

Although this work has attempted to provide a "pan-dialectal" analysis of Kurmanji, the pharyngealized obstruents, attested only in some dialects (Bidlîs, Sêrt), are worth mentioning here because of another apparent ex-

[10] Not merely analogy, by which phonological similarity draws items phonologically closer (Garrett 2015), some discursive similarity, in this case sequence, causes phonological contamination of one item by another.

Table 16: 'Eleven', 'twelve', 'thirteen', 'fifteen', 'sixteen', and 'seventeen'

Proto-Indo-European	Avestan	Kurmanji	Gloss
*oi̯nos-deḱm̥-	aēvan-dasa-	[jãzdæ]	'eleven'
*du̯ō-deḱm̥-	duwa-dasa-	[dãzdæ]	'twelve'
*trei-deḱm̥-	θri-dasa-	[sezdæ]	'thirteen'
*penkwe-deḱm̥-	panca-dasa-	[pãzdæ]	'fifteen'
*su̯eḱs-deḱm̥-	xšvaš-dasa-	[ʃãzdæ]	'sixteen'
*septm̥-deḱm̥-	hapta-dasa	[ħævdæ]	'seventeen'

ception. Consider the item [mæzˤɪm], 'great', whose /z/ → [zˤ] shift would appear in line with the analysis up to this point, both in terms of syllable structure for a pharyngealized vowel (it is preceded by [æ]), and in terms of the labial /m/ on the other side of the vowel. So it would appear that pharyngealized obstruents in surface forms of inherited vocabulary in these dialects simply add several coronal obstruents (/t/, /d/, /s/, /z/, those whose pharyngealized equivalents are attested in Arabic) to /h/ to create a group of consonants on which pharyngealization may "land". To this clean argument I immediately counterpose [zˤæn]-, 'to know' (see Table 1). [zˤæn] ([zɑn] in most dialects) has the coronal and non-flat nasal /n/ instead of a labial or indeed any flat consonant. This could, however, be accounted for due to association with the semantically and phonologically similar Arabic verb /zˤann/, 'to suppose' or 'to reckon'. Arabic verbs are regularly loaned into Kurmanji with an *-in* infinitive form: the infinitive of [zɑn] is [zɑnin]), formally similar to the result of the loaned Arabic /maʃʔ/, 'to walk' > Kurdish [mæʃin].

5 Conclusions

The existence of pharyngeal articulations in Kurmanji might be expected due to the long history of contact and likely bilingualism between speakers of Kurdish and speakers of Semitic languages, particularly Arabic. However, as previously discussed, pharyngeals in inherited vocabulary, being non-etymological, must be accounted for by a different process to simple lexical borrowing.

Thus far, I have put forth several hypotheses about pharyngeals in inherited vocabulary in Kurmanji:

1. Pharyngeals have arisen spontaneously in many syllables containing a low vowel and a labial, or a round mid-high vowel and [ʃ], due to their combined effect on the formant frequencies approximating the effects of pharyngealization.

2. Exceptional cases can be explained in terms of phonological analogy or the influence of contamination with other, already pharyngeal items in the language.

3. Kurmanji pharyngeals occur in permissible contexts for approximants, with the voiceless [ħ] replacing /h/ where it occurs.

These generalizations make predictions which are generally born out in practice in Kurmanji. However, exceptions may still be found which necessitate further study into this pattern. For example, /bɑʃ/ ('good') is a common root that contains a labial consonant and a low vowel, but it has never been reported with a pharyngeal, and speakers of various dialects reject [bæʕʃ] as a possible pronunciation of the word (in their dialect or any other with which they are familiar). Further restrictions may be posited to explain such cases. There are no apparent cases of pharyngeal-voiceless fricative clusters (-VʕS) in the language, including in Arabic loans where they might be expected: [mɑf], 'right(s), entitlement' is generally agreed to derive from Arabic /miʕaːf/, but is never realized as [mæʕf] in Kurmanji.

This paper set out to explain the appearance of pharyngeals in Kurmanji in phonetic and phonological terms. Previous analyses fail to account for the process of incorporation of pharyngeals into the phonology vis-à-vis inherited vocabulary, although this phenomenon is noteworthy. While linguists are willing to assume discrete, compartmentalized phonologies or phonological rules for loan vocabulary and inherited vocabulary (Hall 2013: 238–239, 246–250), the varied and possibly expanding store of inherited items with pharyngeals implies a unified phonology.

An interesting consequence of the analysis has been the necessity of reinterpreting the pharyngeal sounds not as consonants, but as pharyngealized vowels, owing to the apparent constraint on which vowels may serve as nuclei for pharyngealized syllables. My claim is that, synchronically, Kurmanji contains two additional pharyngealized vowel phonemes, /æˤ/ and /ɪˤ/.

Pharyngeals, whatever their origin, are salient in Kurmanji, and speakers of the language (as well as their linguistic neighbors) are acutely aware of this sound, and view it as a differentiating feature between the "Kurdish accent" and the "Turkish accent" or "Persian accent" (but not the "Arab

accent", which may explain their exclusion from the Latin script orthography). Saliency is often ranked as an important criterion in determining the sounds of a language, but without minimal pairs, the contrast between pharyngeal and non-pharyngeal vowels in Kurmanji, however intriguing, would be a case of a "quasi-phonemic contrast" (Hualde 2004; Scobbie & Stuart-Smith 2008). Do minimal or near-minimal pairs exist to contrast pharyngeal and non-pharyngeal vowels in Kurmanji?

One difficulty in locating minimal or near-minimal pairs in the language is that one of the posited pharyngealized vowel phonemes, /ɪˤ/, is relatively rare. Many Arabic words with a high vowel adjacent to a pharyngeal find /æˤ/ forms in Kurmanji, which is perhaps expected given that pharyngeal articulation has an F1 raising effect reminiscent of vowel lowering (Ghazeli 1981). Of the few inherited items which contain /ɪˤ/ (Table 13), none have a non-pharyngeal equivalent in loan vocabulary.

The more common /æˤ/ would seem the more fruitful place to look for minimal pairs. And in spite of the capacity for new items to become pharyngealized, we see the emergence of a small number of consistent minimal and near-minimal pairs with the low pharyngeal vowel, as in (3).

(3) Minimal and near-minimal pairs with the low vowels /æˤ/, /ɑ/, and /æ/
   a. /æˤɾd/ [ʕærd] 'ground', /ɑɾd/ [ɑɾd] 'flour'
   b. /æˤvdɑn/ [ʕævdɑn] 'slaves' (oblique), /æv dɑn/ [ævdɑn] 'gave them'
   c. /hæˤvɑn/ [ħævɑn] 'pieces' (oblique), /hævɑl/ [hævɑl] 'friend'

The rarity of pharyngeal sounds cross-linguistically and their complex evolution in Kurmanji Kurdish makes the language crucial for the study of areal sound patterns. Significant variation exists within the language, including with regard to pharyngeals, making all Kurdish varieties valuable to continued research. This work cannot be delayed, as many dialects of the language are under threat by ongoing assimilation, the result of the repression, migration, and the language policies of the modern states in control of the four parts of Kurdistan, which have followed Kurmanji speakers into the diaspora. This is particularly true for Turkey, with its large Kurmanji-speaking population whose language transmission has been severely threatened by decades of denialism, a generation-long total ban on their language, and whose fundamental rights to education in and propagation of their own language are still unrecognized and impeded to this day.

To conclude, it is hoped that this small study underscores the importance of preserving, documenting, learning, and passing on this geographically, historically, and culturally significant language, and all under-recognized, under-studied, and threatened languages. If a language such as Kurmanji, with its millions of speakers, can lose so much linguistic ground over the past century, and can, in the early 21st century, still have so much linguistic work to be done as this study on one corner of its phonology implies, we can only imagine the state of even smaller and less known languages, within the Iranian language family, within the Middle East region, and around the world.

References

Bird, Brian A. 2007. *Aspects of Yaghnobi grammar*. University of Oregon dissertation.

Blevins, Juliette. 1995. The syllable in phonological theory. In Goldsmith, J. (ed.), *The handbook of phonological theory*, 206–244. Cambridge, MA: Blackwell. [Reprinted in Kreidler, C. W. (ed.). 2001. *Phonology: Critical concepts*. Vol. 3. 75–113. London: Routledge].

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2017. Areal sound patterns. In Hickey, R. (ed.), *The handbook of areal linguistics*, 88–121. Cambridge: Cambridge University Press.

Chyet, Michael L. 2003. *Kurdish-English dictionary: Ferhenga kurmancî-inglîzî*. New Haven/London: Yale University Press.

Cook, Eung-Do. 1993. Chilcotin flattening and autosegmental phonology. *Lingua* 91(2–3). 149–174.

Edelman, Džoy I. & Dodykhudoeva, Leila R. 2009. The Pamir languages. In Windfuhr, G. (ed.), *The Iranian languages*, 773–786. New York/London: Routledge Press.

Esling, John H. 1999. The IPA categories "pharyngeal"and "epiglottal": Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and speech* 42(4). 349–372.

Esling, John H. 2010. Phonetic notation. In Hardcastle, W. J. & Laver, J. & Gibbon, F. E. (eds.), *The handbook of phonetic sciences*, 678–702.

Fortson IV, Benjamin W. 2010. *Indo-European language and culture: An introduction*. 2nd edn. Malden, MA: Wiley-Blackwell.

Fulop, Sean A. & Kari, Ethelbert & Ladefoged, Peter. 1998. An acoustic study of the tongue root contrast in Degema vowels. *Phonetica* 55(1–2). 80–98.

Garrett, Andrew. 2015. Sound change. In Bowern, C. & Evans, B. (eds.), *The Routledge handbook of historical linguistics*, 227–248. London: Routledge.

Ghazeli, Salem. 1981. La coarticulation de l'emphase en arabe. *Arabica* 28(2–3). 251–277.

Haig, Geoffrey. 2018. Northern Kurdish (Kurmanjî). In Haig, Geoffrey & Khan, Geoffrey (eds.), *The languages and linguistics of western Asia. An areal perspective* (The World of Linguistics 6), 106–158. Berlin: De Gruyter Mouton. (`https://www.degruyter.com/view/product/456121`) (Accessed 2019-08-26).

Haig, Geoffrey & Matras, Yaron. 2002. Kurdish linguistics: A brief overview. *STUF – Language typology and universals* 55(1). 3–14.

Haig, Geoffrey & Öpengin, Ergin. 2014. Kurdish: A critical research overview. *Kurdish studies* 2(2) [Special issue: *Kurdish linguistics: Variation and change*]. 99–122.

Haig, Geoffrey & Öpengin, Ergin. 2018. Kurmanji Kurdish in Turkey: Structure, varieties and status. In Bulut, C. (ed.), *Linguistic minorities in Turkey and Turkic-speaking minorities of the peripheries*, 157–229. Wiesbaden: Harrassowitz.

Hall, Kathleen C. 2013. A typology of intermediate phonological relationships. *The linguistic review* 30(2). 215–275.

Hoberman, Robert D. 1985. The phonology of pharyngeals and pharyngealization in Pre-Modern Aramaic. *Journal of the American Oriental Society* 105(2). 221–231.

Holes, Clive. 1995. Community, dialect and urbanization in the Arabic-speaking Middle East. *Bulletin of the School of Oriental and African Studies* 58(2). 270–287.

Holes, Clive. 2004. *Modern Arabic: Structures, functions, and varieties*. Washington DC: Georgetown University Press.

Hualde, José I. 2004. Quasi-phonemic contrasts in Spanish. In Chand, V. & Kelleher, A. & Rodríguez, A. J. & Schmeiser, B. (eds.), *WCCFL 23: Proceedings of the 23rd West Coast Conference on Formal Linguistics*, 374–398. Somerville, MA: Cascadilla Press.

Jakobson, Roman & Fant, Carl G. M. & Halle, Morris. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jastrow, Otto. 1980. Die Dialekte der Arabischen Halbinsel. In Fischer, W. & Jastrow, O. (eds.), *Handbuch der arabischen Dialekte*, 103–122. Wiesbaden: Harrassowitz.

Kahn, Margaret. 1976. *Borrowing and variation in a phonological description of Kurdish*. University of Michigan dissertation.

Karimi-Doustan, Gholamhossein. 2002. The structure of the syllable in the Kurdish language. *Journal of the Faculty of Literature and Humanities of Mashad* 35(1–2). 235–248.

Khromov, Albert L. 1972. *Yagnobskij jazyk*. Moscow: Nauka.

McCarus, Ernest N. 2009. Kurdish. In Windfuhr, G. (ed.), *The Iranian languages*, 587–633. London/New York: Routledge.

Ohala, John J. 1985. Around flat. In Fromkin, V. (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 223–241. London: Academic Press.

Paul, Ludwig. 2008. Kurdish language I: History of the Kurdish language. In *Encyclopaedia iranica online*. (`http://www.iranicaonline.org/articles/kurdish-language-i`) (Accessed 2017-06-26).

Pokorny, Julius. 2007. *Proto-Indo-European etymological dictionary: A revised edition of Julius Pokorny's Indogermanisches etymologisches Wörterbuch*. Published online by Indo-European Language Revival Association. (`https://marciorenato.files.wordpress.com/2012/01/pokorny-julius-proto-indo-european-etymological-dictionary.pdf`) (Accessed 2017-06-29).

Scobbie, James M. & Stuart-Smith, Jane. 2008. Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In Avery, P. & Dresher, B. E. & Rice, K. (eds.), *Contrast in phonology: Theory, perception, acquisition, phonology and phonetics*, vol. 13, 87–113. Berlin: Mouton de Gruyter.

Thackston, Wheeler M. 2006. *Kurmanji Kurdish: A reference grammar with selected readings*. Cambridge, MA: University of Harvard. (`https://sites.fas.harvard.edu/~iranian/Kurmanji/kurmanji_complete.pdf`) (Accessed 2019-07-08).

Thordarson, Fridrik. 1989. Ossetic. In Schmitt, R. (ed.), *Compendium linguarium iranicarum*, 456–479. Wiesbaden: Reichert.

Watson, Janet C. 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Watson, Janet C. 2012. *The structure of Mehri* (Semitica Viva 52). Wiesbaden: Harrassowitz.

Windfuhr, Gernot. 2009. Dialectology and topics. In Windfuhr, G. (ed.), *The Iranian languages*, 5–42. New York/London: Routledge Press.

4
Towards a dialectology of Southern Kurdish: Where to begin?

Sara Belelli

Abstract: This contribution provides an overview of the current state of knowledge on the dialectology of Southern Kurdish (hereafter SK). The introductory paragraphs discuss the concept of SK, survey existing sources and briefly address core issues of terminology. The bulk of the study reviews Fattah's (2000: 9) proposed dialect classification, and complements it with the evaluation of language data from older sources, the author's own research in Kermānshāh Province and other documentation activities recently carried out in the SK-speaking area, sketching possible directions for future research.

1 What is Southern Kurdish?

Despite a growing scholarly interest in SK dialectology, SK vernaculars are still among the least documented contemporary Iranian languages and suffer from having too long been relegated to the fringes of linguistic research.[1]

SK can be defined as a bundle of closely related vernaculars, spoken as mother tongue by a minority of the Kurdish-speaking population,[2] mostly living at the southernmost periphery of the core Kurdish-speaking region of the Middle East.

The area where SK is predominantly spoken is rather wide and almost completely included within the borders of present-day Iran: As shown in Figure 1, it stretches (north to south) from the county of Qorve (Kordestān Province),

[1]I wish to thank the anonymous reviewer, later revealed as Erik Anonby, for contributing with punctual suggestions and criticism to the improvement of a previous version of this paper. Of course, I bear responsibility for all the remaining errors and shortcomings.

[2]Figures close to three million people have been proposed (Fattah 2000: 4), but as the reviewer pointed out these include Laki speakers and are likely to be overstated.

to the counties of Ābdānān and Dehlorān (Ilām Province), in the north-central part of the Zagros mountain range. The SK domain also includes a narrow stretch of land on the western side of the Iran-Iraq border and reaches, to the east, the Iranian county of Tuyserkān (Hamadān Province). The SK-speaking enclave of Bijār, located in a mainly Central Kurdish (hereafter CK) linguistic milieu,[3] represents the northernmost outpost of this dialect group.

The region concerning us here is characterized by intricate linguistic geography, prevalent multilingualism and extensive language contact, due to the historical presence, alongside the majority SK-speaking population, of communities speaking other Iranian (i.e. CK, Gorāni, Laki, NLori, Persian[4]) and non-Iranian (i.e. Neo-Aramaic, Turkic and Arabic[5]) languages and dialects. Language variation often parallels the rifts traced by religious and/or ethnic affiliations, which need to be carefully looked into when accounting for the complex distribution and multifaceted interaction between different vernaculars within the community of speakers.

SK has never been subjected to processes of language standardization and planning, nor does it boast a long written literary history.[6] These circumstances have hindered the emergence of a prestigious normative supralect, ensuring the permanence of a globally high level of dialectal heterogeneity. Even today, SK varieties are only rarely written[7] and their use has remained

[3] CK (i.e. Sine'i varieties, cf. De Morgan 1904; Christensen & Barr 1939: 198–234) is the majority language west of the SK *enclave* of Bijār, cf. the language map of Kordestān Province in Anonby & Taheri-Ardali, et al. (2015–2019).

[4] In addition to Sine'i (cf. Footnote 3), so-called 'Southern Jāfi' varieties of CK (Fattah 2000: 3) are spoken in south-west Kordestān Province, north-west Kermānshāh Province and in adjoining areas of Iraq. Gorāni/Hawrāmi dialects are mainly spoken in the Awrāmān region, on the north-western tip of Kermānshāh Province, and adjoining areas of Iraq. Related Gorāni vernaculars are also found in other localities of Kermānshāh Province (cf. Mahmoudveysi & Bailey 2013: 2). Laki and NLori varieties are spoken further south-east of the SK-speaking area, in Ilām Province, Kermānshāh Province and northern Lorestān (cf. Figure 1).

[5] Turkic varieties are spoken around Bijār and Qorve and in a few other spots within the SK-speaking area (cf. Fattah 2000: 2–3, 5, 17–19; Anonby et al. in this volume). After a mass migration outside Iran during the 20th century, only a few families speaking Jewish NENA dialects are still found in the region (cf. Anonby et al. in this volume; Khan 2009: 5-11; 2011). SK is also in contact with the Arabic vernaculars of eastern Iraq and with a pocket of Arabic speakers in north-east Ilām Province, cf. the language maps of Ilām Province in Aliakbari et al. (2014) and Anonby & Taheri-Ardali, et al. (2015–2019).

[6] The main linguistic means for oral and written transmission of literary works in the SK-speaking region has historically been literary Gorāni (cf. Kreyenbroek & Chamanara 2013). Currently, the role of regional dominant languages has been taken on by Persian (in Iran) and Arabic and Sorani Kurdish (in Iraq), by virtue of their status as official languages of education and administration.

[7] The few attempts at writing SK either apply the Sorani writing norm or involve a modified

largely confined to everyday oral communication, especially in the domestic sphere.

Figure 1: Respective distribution of SK, Laki and Northern Lori (hereafter NLori) dialects, according to Fattah (2000: iv)

Despite their surface complexity, however, SK dialects form a rather compact dialect continuum, unified by a fair degree of mutual intelligibility based on shared phonological, morphosyntactic and lexical features. To my knowledge, no specific study on mutual intelligibility between any two SK varieties has ever been carried out, but personal observations and other hints from

Arabo-Persian orthography, which can vary considerably from writer to writer (personal observation).

the existing literature Aliakbari et al. (2014: 7-8) suggest that dialect variation rarely impedes face-to-face interactions.

The tricky question of the linguistic affiliation of Laki – an associated variety spoken further south-east of the SK-speaking area[8] – is still a debated issue and will be briefly touched upon at the end of this paper. For our purposes, the narrowest possible definition of SK is adopted and Laki varieties will be considered as forming a separate cluster of Northwest Iranian varieties. The controversial question whether or not Laki is part of the Kurdish language continuum will also be left unanswered for the time being.

1.1 Sources

The scarcity of detailed information on most SK varieties and the variable quality and types of the existing data make a complete and reliable account of the dialect situation a long-awaited desideratum. Indeed, these varieties pose a genuine challenge to dialectologists, being both extremely diverse and severely under-documented.

In spite of the relevance of this group of vernaculars for Kurdish and Iranian linguistics, the number of scholarly works dealing with SK is inexplicably low. At present, the natural starting point is the monograph published by Fattah (2000) almost two decades ago, which in spite of all shortcomings still represents the most comprehensive collection of information on individual SK vernaculars, as well as an initial attempt at sketching a group-internal classification (cf. Section 2). The language data contained therein, however, remains for the most part unverified.[9]

A few works predating Fattah's monograph (e.g. Blau 1989; Christensen & Barr 1939; De Morgan 1904; Querry 1896) also provide interesting data for comparison on individual varieties (i.e. the SK of Bijār, a not better specified 'Kermānshāhi' dialect and the Badre'i of Baghdad).

A significant amount of descriptive work on other SK vernaculars has appeared in Iran (e.g. Karimi-Doostan 1380/2001 on the SK of Badre; Morādi 1394/2015 on several verieties of Kermānshāh Province; Aliyāri Bābolqāni

[8] Figures close to a million speakers have been reported for Laki (Fattah 2000: 4; Izady 1992: 175), but I consider that these are exaggerated, perhaps including Laki-Kermānshāhi (cf. Section 2) and Khezeli speakers (cf. Aliakbari et al. 2014: 7-8) in the count (Erik Anonby, p.c.).

[9] In this respect, my own research (Belelli 2016) has confirmed the overall accuracy of the information reported in Fattah (2000) for the vernacular of Harsin. With due caution, this can reassure those using that source on the reasonable correctness of data from other SK varieties contained therein.

1396/2017 on the varieties of Gilān-e Gharb and Gahvāre). As for dictionaries, Darvishyān (1375/1996), Jaliliyān (1385/2006) and Karimpour (1382/2003) can be quoted among the best-known references.

Documentation of SK is recently being revived by newly founded research groups, which have already achieved important results and augur well for the progress of research in this field.[10]

1.2 Issues of terminology

Due to a great deal of confusion surrounding many labels applied to SK vernaculars, a clarification of terminology is an appropriate point of departure for any study dealing with the subject.

Looking at endonyms, we observe that native speakers usually refer to individual SK varieties simply as 'Kurdish' (SK *kordī*, *kwirdī*, etc.). If they need to stress the distinctness of their own or other vernaculars *vis-à-vis* neighboring groups, they apply a reflexive pronoun (e.g. *kordī xomān*, *kwirdī wižmān*, etc. 'our own Kurdish') or a tribal-geographic specifier (e.g. *kordī kalhorī* 'Kurdish spoken by Kalhors', *kordī kirmāšānī* 'Kurdish spoken in Kermānshāh city/area') to the generic language label. I could observe that at least a section of Laki-Kermānshāhi speakers (cf. Section 2) refer to their dialects also as *lakī*, although they seem quite categorical in stressing their distinctness from the communities of Kermānshāh Province that speak varieties akin to the Laki of northern Lorestān and eastern Ilām (Belelli 2016: §1.4).

In the literature, SK dialects have been collectively referred to in various ways. Terms originally having a more restricted semantics, usually connected to (historical and/or contemporary) place names and ethnic groups (e.g. Kalhori,[11] Kermānshāhi,[12] Feyli,[13] Pa(h)lawāni) have been used by popular schol-

[10] The project *Documentation of Gorani, an endangered language of West Iran* (cf. Mahmoudveysi et al. 2012; Mahmoudveysi & Bailey 2013) has restituted, as a side result, two samples (i.e. *laki_conv_1* and *laki_conv_2* `https://hdl.handle.net/1839/00-0000-0000-0018-03DC-B@view`, retrieved June 2018) of varieties pertaining to the Laki-Kermānshāhi subgroup (cf. Section 2). SK language samples and other kind of linguistic and bibliographic information is also being collected by the teams of Anonby & Taheri-Ardali, et al. (2015–2019) and Matras et al. (2016).

[11] This choice might be motivated by the wide diffusion of Kalhori dialects (cf. Section 2) and/or by their higher prestige in comparison to other SK varieties.

[12] Labels such as *Kermānshāhi, Ilāmi*, etc. might be confusing in that the geographic/administrative categories on which they are based can refer to entities of variable size (city, county, province, etc.) at the same time.

[13] According to Fattah (2000: 70–74), only some SK speakers of Baghdad used the term 'Feyli' as self-denomination. Historically, the term denoted the principality of Lor-e Kuchek ('Lorestān-e Feyli', i.e. current Ilām and Lorestān Provinces) and was only later extended to qualify the

ars and academics alike as cover terms for SK. I maintain, however, that the use of 'Southern Kurdish' as collective language label should be favored, being already common in scholarly works and less ethnically and geohistorically connotated than most available options.

As neutral as it may be, however, this label is not entirely exempt from ambiguity: since Iraqi Kurdistan is often referred to as 'Southern Kurdistan' (CK *kurdistānī bāšūr*), 'Southern Kurdish' sometimes denotes the varieties of CK (Sorani) spoken there. For this reason, we could consider the labels 'Southeastern Kurdish' (Schmitt 2000: 76–77, from the common designation of Iranian Kurdistan as 'Eastern Kurdistan', CK *kurdistānī řōžhałāt/xōrhałāt*) or 'Southern Iranian Kurdish', common in Iraqi Kurdistan, two viable alternatives for further disambiguation.

2 Dialect distribution and Fattah's (2000) preliminary sub-grouping

Fattah (2000) was probably the first to sketch a classification of SK vernaculars, identifying seven dialect subgroups (listed below from north to south, cf. Figure 2):[14]

1. **Bijāri:** also known as Garrusi,[15] is spoken in the county of Bijār (former Garrus, Kordestān Province).

2. **Kolyā'i:** includes the varieties spoken in the district of Chahārduliye Gharbi (Qorve County, Kordestān Province), in northern and eastern Kermānshāh Province (from the Poshtdarband rural district up to Kāmyārān, in Dinavar district and Khodābandelu rural district). The SK dialects spread in Asadābād, Hamadān, Tuyserkān and Malāyer Counties (Hamadān Province) are also classified in this subgroup.

3. **Laki-Kermānshāhi:** includes the so-called 'Pāyravand' vernaculars (Fattah 2000: 23), spoken in the rural Districts of Dorudfarāmān and Miyāndarband (Kermānshāh County), and most varieties spoken in the Sahne

people inhabiting this region and the vernaculars they speak (i.e. SK and NLori dialects). Its inherent ambiguity, however, makes it a rather infelicitous choice as a language label.

[14] It is worth pointing out that this study appeared quite late in comparison to MacKenzie's (1961a, 1962) major work on regional variation in Kurdish, which did not treat the SK situation in any detail.

[15] cf. Christensen & Barr (1939: 291–331); De Morgan (1904) and Querry (1896)

and Harsin Counties of Kermānshāh Province (except for the rural District of Chamchamāl, where dialects akin to the Kalhori-Sanjābi-Zangane subgroup are mainly spoken).

4. **Kalhori-Sanjābi-Zangane:** includes the varieties spoken by a section of ethnic Sanjābi (roughly located west of Kermānshāh city and east of Qasr-e Shirin, in Kermānshāh Province), the varieties spread in the territories of the Kalhor people (between Eslāmābād-e Gharb in the north and Eyvān in the south, in Kermānshāh and Ilām Provinces) and the vernaculars of the Zangane people of the Harasam rural district (Kermānshāh Province). Varieties akin to Kalhori are also spoken immediately across the border in Iraq, between Khānaqin and the town of Zurbātiya (northeast of Badra). Fattah (2000: 27–28) also seems to ascribe the dialects spoken by the Arkavāz people (between the cities of Eyvān and Ilām, in Ilām Province), and those spoken in the counties of Sirvān and Chardāvol (Ilām Province) to this group.

5. **Malekshāhi:**[16] includes the dialects spoken in the region to the north and west of Badre District (Ilām Province), inhabited by ethnic Malekshāhi and Mishkhās. Varieties pertaining to this group are also spoken in the Sālehābād district (Mehrān County, Ilām Province), around Zurbātiya (in Iraq) and by a section of the SK-speaking population of Baghdad.

6. **Badre'i:** spoken in Badre District (Darre Shahr County, Ilām Province), as well as by a section of the SK-speaking population of Baghdad.[17]

7. **Kordali:** or Ābdānāni (Aliakbari et al. 2014) spoken at the southern periphery of the SK-speaking area, in the counties of Dehlorān and Ābdānān (Ilām Province) occupied by ethnic Kordali, and adjacent areas of Iraq.

[16] This cluster roughly combines the 'Ilāmi' and 'Malekshāhi' subgroups in Aliakbari et al. (2014: 7). The reviewer correctly pointed out that Fattah's label looks rather odd, considering that ethnic Malekshāhi are neither the larger, nor the most representative community speaking varieties related to this subgroup.

[17] cf. Blau (1989)

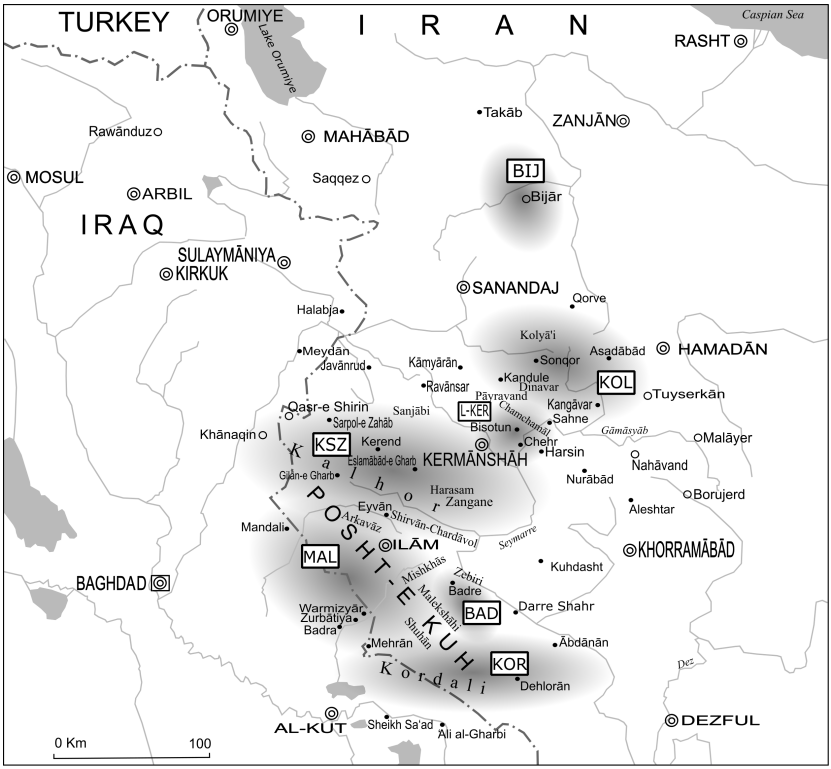

Figure 2: Approximate distribution of SK dialect subgroups, based on the evaluation of the sources outlined in Section 1.1 (BIJ: Bijāri; KOL: Kolyā'i; L-KER: Laki-Kermānshāhi; KSZ: Kalhori-Sanjābi-Zangane; MAL: Malekshāhi; BAD: Badre'i; KOR: Kordali).

To summarize, leaving the enclave of Bijār aside, the dialects pertaining to the second, third and fourth groups are mainly spoken in Kermānshāh Province and adjoining areas of Kordestān, Hamadān, Ilām and Iraq, while the dialects pertaining to the last three dialect groups are uniformly spoken in Ilām Province and adjoining areas of Iraq (i.e. historical Posht-e Kuh). The latter correspond to the 'Ilāmi (Feyli) dialect group' in Aliakbari et al. (2014: 7–8), although these authors suggest a different internal classification.

Unfortunately, Fattah's material is unevenly arranged and the author fails to specify the features setting out each subgroup from the others. From the

chosen denominations it seems clear that his dialect subgroups, despite the terminology used, should not be considered simply as *geolects*,[18] but are first and foremost *ethnolects*.[19] One can suppose that their distribution has become more sharply defined in space only in the aftermath of the mass sedentarization of pastoral-nomadic tribes of West Iran, particularly after the first two decades of the 20th century (cf. Potts 2016: 428–429).

Another necessary proviso is that any assertion of dialectal uniformity, at any level of linguistic or spatial analysis (e.g. within a single dialect subgroup or even within a single village), is an inevitable oversimplification and an ultimate artifact in a region characterized by ongoing, widespread human mobility and frequent displacement of population groups.[20] At this stage of what is known, any representation of SK dialects in physical space cannot account for the complex intersections and frequent overlapping of dialect areas, or the existence of transitional idioms and 'mixed' urban dialects (e.g. the Persian/SK creole of Kermānshāh city, cf. Guizzo 2007).

3 Dialect features

The following paragraphs contain a selection of dialect features which may be relevant in identifying clusters of SK vernaculars, as assembled from the sources outlined in Section 1.1. They are divided into phonetics and phonology (Section 3.1), morphosyntax (Section 3.2) and lexicon (Section 3.3). No quantitative evaluation of the data has been carried out and no figures concerning relative distance between different SK dialects or subgroups are currently available.

3.1 Phonetics and phonology

In order to identify regional patterns of variation, the consonant and vowel inventories of SK dialects have been tentatively represented in a unified table (cf. Table 1).

[18] A more or less uniform group of varieties spread over an identifiable geographic area.

[19] A variety associated to a specific ethno-cultural group, generally a tribe or a section of it.

[20] Notwithstanding the decline of pastoral nomadism in west Iran, population movements have continued until the present in the form of migratory flows from rural areas towards major urban centers.

Table 1: Consonant and vowel inventories of SK dialects (including peripheral sounds)[21]

	Labial	Dental-Alveolar	(Pre)-palatal	Velar	Uvular	Pharyngeal	Glottal
Voiceless Stops/Affricates	p	t	č	k	q		
Voiced Stops/Affricates	b	d	j (gʲ)*	g			(ʔ)
Voiceless Fricatives	f	s (ṣ)*	š	x		(ḥ)*	
Voiced Fricatives	(v)*	z (ẓ)*	ž	(ġ)*		(ʕ)*	h
Nasals	m	n		(ŋ)*			
Laterals		l ɬ*	(ʎ)*				
Vibrants		r ř					
Semivowel	w (ẅ)*		y				

	Front	Central	Back
Close	ī ü*		u
Mid	(e) (ö)*	i	(o)
Open	a		ā

Consonants: The consonant inventory of SK dialects is rather uniform, with only a few sounds (mostly peripheral in the system) being confined to specific dialect subgroups or single varieties:

Phonemic /v/ [v] is more or less restricted to L-KER and some KOR vernacu-

lars, a feature perhaps attributable to proximity to Laki and NLori (cf. Lazard 1992: 216; MacKinnon 2002: 106). In L-KER and some KOR dialects, /v/ [v] parallels cSK /w/[w], especially in word-initial and intervocalic positions and as common outcome of lenition of an original intervocalic [b]: e.g. *vark* vs. cSK *wark* 'lamb'; *vitin* vs. cSK *watin*, *witin* 'say'; *mīva* (≠ L-KER of Sahne *mīwa*) vs. cSK *mīwa*; *xavar* vs. cSK *xawar* (≠SK of Kermānshāh *xabar*). This sound is virtually unknown to other SK varieties, with isolated exceptions (in word-final position or as allophone of /b/ [b] and /f/ [f] before /d/ [d]), e.g. SK of Qasr-e Shirin *dīv* 'demon' (Fattah 2000: 94); BAD *avdāł* 'monk' (Blau 1989: 44 ff.).

Dark /ɬ/ [ɬ] appears to be phonemic in virtually all SK varieties, except for the varieties of Kermānshāh city and Mandali, where it is regularly replaced by clear /l/ [l]: e.g. *bāl* vs. cSK *bāɬ* 'arm'; *pyāla* vs. cSK *pyāɬa* 'cup' (Fattah 2000: 98–99). A similar distribution applies to the velar nasal [ŋ],[22] e.g. Ker., Man. *tang* vs. cSK *taŋ* 'narrow, tight', equally rare in the SK of Qasr-e Shirin and Xānaqin (Fattah 2000: 104). Another interesting trait concerning laterals is the frequent palatalization of /l/ to [ʎ] in some L-KER dialects, e.g. L-KER of Harsin *lā* [ʎɑ:] '(be)side'; *kwil* [kwəʎ] 'all' (Belelli 2016: §2.1.3), a feature also observed in Laki and NLori dialects (Anonby 2004–2005: 16; MacKinnon 2002: 107).

A palatalized realization of /g/ [g] in word-medial (mainly intervocalic) and final position is typical of the dialects spoken in Iraq and border areas (especially MAL, but also some Kalhori and Sanjābi dialects). In MAL varieties, g^j [ɟ] generally parallels cSK word-final /g/ or /k/, e.g. *kilig*^{j} vs. cSK *kilik*, *kilig* 'finger'. In BIJ and northernmost KOL dialects (e.g. Qorve) [ɟ] results from lenition of an original intervocalic/postvocalic *d*,[23] yielding [y] in other SK dialects, e.g. *ziyāg*^{j} vs. cSK *ziyā(y)*, *zīād* 'much, many'; *āg*^{j}*am*, *āg*^{j}*im* vs. cSK *āyam*, *āyim* 'person' (Fattah 2000: 100–101, 135 ff.).

SK vernaculars close to the border with Iraq (i.e. MAL, some KOR and Kalhori vernaculars spoken along the border) see the presence of emphatic [sˤ] [zˤ]

[21] Peripheral sounds (i.e. non-integrated, used in borrowed items only, having low frequency and/or uncertain phonemic status) are marked by parentheses, while sounds restricted to certain varieties or groups of varieties (Fattah's 'non-generalized' sounds) are marked by *. In examples, the abbreviation 'cSK' precedes the forms common to most SK dialects.

[22] Minimal pairs contrasting /ɬ/ (never found word-initially) point to its phonemic status with incomplete distribution, e.g. *kol* 'rounded, smoothed' / *koɬ* 'short' (Fattah 2000: 98), *bār* 'bring!' / *bāɬ* 'wing'. [ŋ] (never found word-initially) can be seen as predictable allophone of /n/ before homorganic consonants (/k/ and /g/), but contrasting pairs do exist, e.g. *taŋ* 'narrow, tight' / *tan* 'body'; *saŋ* 'stone' / *sag* 'dog'. We mention here also the incomplete distribution of the flap /r/, regularly replaced by trilled /ř/ word-initially.

[23] This phenomenon is generally referred to as 'Zagros-d' (McCarus 2009: 591).

and pharyngeal [ħ] as common counterparts of corresponding non-emphatic sounds in both inherited and borrowed words, pointing to a stronger influence, whether direct or indirect, of Arabic, e.g. *ṣuẓānin*, *ṣoẓānin* vs. cSK *suzānin* 'burn', *gonāḥ* vs. cSK *gonā(h)*, *gwinā(h)* 'sin' (Fattah 2000: 96–97, 107). The retention of Arabic *ᶜayn* in loans is also common in these dialects, e.g. *ʕāqiɬ* vs. cSK *āqiɬ* 'wise' and the insertion of a word initial pharyngeal often extends to non-Arabic items as well, e.g. KOR *ʕas(i)p* vs. cSK *asp* 'horse' (Fattah 2000: 106–107).

Many varieties of Ilām and adjacent areas of Iraq tend to preserve [ɣ] in Arabo-Persian and Turkic loans vs. cSK [q] or [χ], e.g. *āġā* vs. cSK *āqā*, *āxā*. The overall tendency in the varieties not retaining [ɣ] is to favor a realization as /q/ [q] towards the north (e.g. L-KER; KOL; but ≠ BIJ, aligning with southern dialects) and as /x/ [χ] towards the south of the SK-speaking area (e.g. some Kalhori, Zangane and MAL dialects, BAD, cf. Fattah 2000: 215–216).

Vowels: Fattah (2000: 75) states that SK vowel system manifests more regional variation than it is attested for consonants.

The vowel /a/ [a] is very unstable and a certain degree of free variation with the central vowel /i/ [ə] in unstressed syllables is ubiquitous across SK.

Front-rounded vowels /ü/ [y, y:] and /ö/ [ø] are not common to all SK dialects: The first is absent in most MAL and KOR varieties, having /ī/ [i, i:] in comparable contexts, e.g. *dīr* vs. cSK *dür* 'far', *šī* vs. cSK *šü* 'husband'. In the dialects lacking /ü/, the labiopalatal approximant [ɥ] (represented as ẅ in Fattah 2000: 110) is also missing.[24] The sound [ø] has very low frequency and uncertain phonemic status. It is virtually restricted to L-KER, KOL and KOR dialects, where it most often results from the fronting and rounding of /a/ before /w/, e.g. *šöw* vs. cSK *šaw* 'night' or from the dropping of an original *h*, e.g. *nö* vs. cSK *no(h)*, *nu* 'nine'.

Mid vowels /o/ [o, o:] and /e/ [e, e:] pattern as peripheral: the first seems to be found in all dialects, although it is often diphthongized to [wə], e.g. *kwiř koř* 'boy', or replaced by /i/, e.g. *jift* ~ *joft* 'pair', particularly (but not exclusively) in Kalhori dialects (Fattah 2000: 117 ff.). In KOR vernaculars, possibly under NLori influence (cf. MacKinnon 2002: 109), cSK /ā/ [ɑ, ɑ:] is raised to [o, o:] before nasals (with concomitant consonant dropping in the case of /n/) or in the group /āy/, e.g. *gyo(n)* vs. cSK *gyān* 'life, soul', *čoy* vs. cSK *čāy* 'tea'. /e/ [e, e:] and its slightly diphthongized allophone [ej] occur in loan-

[24] I provisionally take this sound as an allophonic variant of /w/, produced by a series of coarticulatory processes involving an original group /wī/ following /a/ (in turn fronted and rounded to [ø]), e.g. **tawīla* >*töẅla* [tøɥ'la] 'stable, cattle shed'. However, as the reviewer pointed out, it might be phonemic in some SK dialects.

words or result from the contraction of /ay/, e.g. *me(h)mān* 'guest', *kaywānu* vs. *kewānu* 'old woman'. Their status and the patterns of alternation with /ī/ are still largely unclear.

As already seen, (historical-phonemic and allophonic) lenition of intervocalic and postvocalic voiced stops /b, d, g/ and word-initial and word-final /b/ is characteristic of SK, although slightly less prevalent in urban centers (perhaps under the pressure of standard Persian pronunciations). Morphophonemic lenition also occurs when present and past verbal stems beginning with /b, d/ are preceded by indicative, subjunctive or negative prefixes, a complex phenomenon[25] that seems to manifest different regional outcomes (i.e. from full maintenance of the consonantal sounds, e.g. BAD *adan* 'they give', to their complete dropping, e.g. L-KER of Harsin *men* 'they give', with various intermediate stages, e.g. KOL *ayan*, BIJ *dig^{j}an* 'they give', cf. Fattah 2000: 408 ff.). At a very first glance, it seems less prevalent in urban dialects (likely influenced by Persian, which does not exhibit this lenition) and in most varieties of Ilām and adjacent areas of Iraq, but more research is needed before a clearer judgement can be made.

SK varieties show more variation than one may expect also in the outcome of historical sound changes considered relevant for locating Kurdish in the context of West Iranian (cf. Fattah 2000: 152–164; MacKenzie 1961b), with frequent shifts towards typical Southwest Iranian developments instead of expected Kurdish ones, e.g. 'eye': most L-KER, MAL *čam, čyam* vs. BIJ, KOL, KSZ, BAD *čaw, čāw*, KOR *čow*; 'groom': BIJ, L-KER, MAL, KOR *zāmā*, KOR *zomā, domā* vs. KOL, KSZ, BAD *zāwā*; 'deer': cSK *āhu, āhü, āhī* vs. KOL, Kalhori *ās(i)k, ās(i)g, āsu*; 'life': cSK *zin(d)agī, zinay* alongside *žiyān, gyān*. This question cannot be further pursued here, but deserves to be carefully addressed.

3.2 Morphosyntax

Variation is observed in virtually all aspects of SK morphosyntax and is often too slight and pervasive to allow the identification of regional patterns.

Morphology: Looking at nominal morphology, differences are mainly found in the form, rather than the type and function of SK morphemes. The cSK indefinite marker is *-ī(k), -ī(g)* (and variants: Fattah 2000: 241), with the final consonant regularly dropped word-finally. The only appreciable difference

[25] The phenomenon is not restricted to SK. Similar developments are attested, for instance, in the Gorani of Gawraju (see Mahmoudveysi et al. 2012: 20–22), as well as in some varieties of CK (see MacKenzie 1961a: 3–4, 19–20) and NK, particularly with stem-initial b- (see Öpengin & Haig 2014: 157–158).

concerns the vernacular of Kermānshāh city and some KOL, MAL and BAD dialects, allowing the realization of the final consonant also in absence of further suffixation (= CK).

Two forms of the definite marker can express definiteness in SK: *-aka, -aga* (and variants) and *-a* (taking the form *-(a)ka, -(a)ga* only after /ā/). According to Fattah (2000: 246), their distribution allows for identification of a northern group (i.e. KOL) using exclusively *-aka, -aga*, and a southern group (i.e. most MAL, but also L-KER ≠ Sahne) favoring *-a*. Other SK vernaculars allow variation between the two, with an apparent preference for the first type towards the north and for the second towards the south of the SK-speaking area.

The plural/collective suffix *-ayl, -(y)al, -el, -gal* (and variants, Fattah 2000: 248) is the common SK plural marker. Only a section of BIJ speakers use the suffix *-ān* as default plural morpheme, probably under the influence of neighboring CK dialects. In other SK vernaculars the type *-ān* is restricted to a few fixed forms and borrowed invariable plural/collective terms (e.g. L-KER of Harsin *atrāfiān* 'entourage, courtiers', ← Pers.) or used with specialized functions (adverbial locative/temporal, e.g. Kalhori *xāłuān* 'at maternal uncles' (house)' (Fattah 2000: 250 ff.).

On the other hand, the suffix *-ān* is widely involved in the formation of definite plurals, a feature that Fattah (2000: 253) identifies as one of the defining elements of geolinguistic classification in this region. Indeed, the type *'-agān'* is used in most BIJ, KOL and KSZ vernaculars, while the type *'-ala(ga)'* is found in L-KER (≠ Sahne) and in most BAD, MAL and KOR dialects.

The use of the cSK *ezāfe* morpheme *=i* manifests considerable variation (Fattah 2000: 264–265). Despite its presence in virtually all SK dialects, only the northernmost varieties seem to realize it regularly, while L-KER and most varieties of Posht-e Kuh (i.e. BAD, MAL, KOR) favor simple juxtaposition.

North/south variation seems to be attested also in other aspects of SK grammar, e.g. the morphosyntactic behavior of numerals (Fattah 2000: 300): in northern subgroups (i.e. some BIJ, KOL, KSZ dialects) a definite noun phrase determined by a numeral can occur in the plural, while the singular is preferred in the rest of SK. A north/south split is also adpositions (Fattah 2000: 583 ff.), with the L-KER dialects standing out from cSK for their use of typical Laki items (cf. Belelli 2016: §2.17; Fattah 2000: 608).

Personal pronouns and demonstratives: Pronominal forms (Fattah 2000: 275 ff.) are nearly identical in all subgroups, but vary in some ways in their phonological form. For full pronouns, the most evident differences concern the 2PL *(h)oma* of L-KER (= Laki vs. cSK *īwa(n), (h)ua, üa, eva*) and the 3PL forms *ownī, öwnī* of some KOR vernaculars (vs. cSK *(a)wān(a), avāna* and variants). Singu-

lar pronominal clitics are rather uniform (cSK *=im, =id/=it, =ī/=e*). The most appreciable difference concerns plural forms, with L-KER and few other dialects having *=mān, =dān/=tān, =yān* vs. cSK *=imān, =idān/=itān, =iyān*[26] (Fattah 2000: 280 ff.). The L-KER subgroup also diverges in respect to reflexive pronouns, having a form *wiž* (= Laki) vs. cSK *xwa-, xo-* (and variants, Fattah 2000: 291).

Demonstrative adjectives are rather uniform, with slight differences in phonological form (*ī/ e/ ay/ (h)āy* 'this, these'; *a/ aw/ (h)āw* 'that, those', Fattah 2000: 314 ff.). More outstanding is the presence, in the majority of SK ≠ most L-KER and KOL vernaculars, of an animacy distinction affecting the form of demonstrative pronouns (Fattah 2000: 317 ff.). The L-KER subgroup is also unique in allowing the use of the plural suffix *-al* to form plural demonstrative pronouns (i.e. *yānala* 'these', *awānala* 'those').

Verbal morphology: Beyond the unifying feature of a generalized nominative-accusative alignment in SK, there seems to be considerable variation in verbal morphology and TAM systems of different subgroups and single SK varieties. Particularly illustrative are the forms of the prefixes intervening in the formation of the indicative present and imperfect tenses (Fattah 2000: 371 ff.): BIJ *di-, a-*; KOL *a-*; L-KER *(=a) ma-* (= Laki, but ≠ Sahne *a-*); BAD *a-*; KOR *di-*. Most Kalhori and MAL vernaculars build the present without an overt marker and contextually lack a morphologically formed Imperfect tense or build it by placing a morpheme *-yā-* between the past stem and personal endings (Fattah 2000: 375 ff.).

Other differences, all deserving closer scrutiny, concern the conjugation of the Perfect (BIJ adding a conjugated present copula to the past participle vs. cSK combining the past stem/past participle with a verbal agreement suffix and a 3SG invariant copula, Fattah 2000: 382); the choice of Preterite vs. Perfect as preferential past tense (Fattah 2000: 374–375); the choice of auxiliaries for the progressive periphrasis (some MAL, BAD and most dialects of Iraq using *nīštin* vs. cSK *dāštin*, Fattah 2000: 504–505; the suffix used for Past Participles (KOL, L-KER, most KSZ *-ī/-e*; BIJ, MAL, BAD *-īg/-īg*^{j}, KOR *-a*); the forms of verbal endings and the Present clitic Copula (Fattah 2000: 465 ff., 514 ff.); the form and placement of preverbs and postverbs (Fattah 2000: 433).

[26] These forms refer to all contexts except following /a/ and /ā/, when SK dialects having the forms with /i/ also drop the vocalic element.

3.3 Lexicon

A great deal of lexical variation is observable throughout SK, even though the impact of such diversity on mutual intelligibility is somehow mitigated by a widespread knowledge among speakers of equivalents pertaining to SK vernaculars different from their own.[27] Overall, the most divergent dialects seem to be those of the L-KER and KOR subgroups, on one side, and of the KOL and BIJ subgroups, on the other.

L-KER dialects share additional lexical items with Laki (e.g. *lam* vs. cSK *zik* 'belly'; *pit* vs. cSK *lüt* 'nose'; *āyl* vs. cSK *zāru, mināł* 'child'; *gwijar* vs. cSK *büčik* and variants 'small'; *kałiŋ* vs. cSK *gawrā* 'big'; *āyštin* vs. cSK *xistin* 'throw'). KOR varieties are heavily influenced by neighboring NLori vernaculars, with shared forms such as *iškam* 'belly'; *bača* 'child', *mas, gap* 'big'; *(ʕ)aftow* 'sun'; *hama* 'all'. To the north of the SK dialect area, BIJ and KOL have additional items in common with neighboring CK dialects, e.g. *hilka* vs. cSK *xā, xāya* 'egg'; *bayāni* vs. cSK *so(b), šöso* and variants 'morning, tomorrow'; *kotān* vs. cSK *dān* 'beat'.

The dialects spoken across the Iraqi border show a bunch of additional Arabic borrowings, mostly attested outside the core vocabulary: e.g. *tallāja/ sallāja* vs. cSK *yaxčāł* (= Pers.) 'fridge'; *šat(t)* vs. cSK *ju* 'river' (= Pers. *ju* 'channel'); *ʕišat* vs. cSK *zin(d)agī, žiyān* and variants.

4 Summing up

Without any intention to reach immediate conclusions or provide final solutions to classificational problems, I believe the following picture can provide a useful starting point for future work dealing with SK dialectology.

Firstly, it seems obvious that the seven subgroups identified by Fattah (2000) cannot be treated as clear-cut, sharply-defined bundles of dialects. They show considerable internal variation and in some cases seem to include dialects manifesting highly diverging features. Some varieties cannot be easily ascribed to any specific subgroup, but are transitional between subgroups[28]

[27] An example of such adaptability is found in the recording titled laki_conv_1 [2:08-2:14] from the DoBeS Archive (cf. Footnote 10).

[28] According to my assessment of Fattah's 2000 data, the Laki-Kermānshāhi variety of Sahne often aligns with neighboring Kolyā'i dialects. Zebiri/Zoheyri (SK *zürī*, also Southern Sirvāni) is transitional between the Kalhori-Sanjābi-Zangane and Badre'i subgroups. The dialects of Zurbatiya and Wārmizyār, on one side, and of Xānaqin, on the other, also seem to diverge quite a lot from most dialects pertaining to their respective dialect subgroups (i.e. Malekshāhi and Kalhori-Sanjābi-Zangane).

or between these and neighboring languages/dialect clusters.[29]

Variation in SK is often related to non-linguistic factors, e.g. religious and ethnic affiliations, historically sedentary vs. mobile lifestyles, or urban vs. rural contexts. Dialect blending is particularly common in major urban centers (e.g. Kermānshāh, Qasr-e Shirin, Ilām), where the constant interaction of inhabitants of different origins and linguistic backgrounds facilitates contact and convergence over time.

Language data, as Fattah (2000) constantly underlines, hint at the existence of a north/south split, in some way distinguishing the subgroups centered in Kermānshāh Province (i.e. KOL, KSZ) from those spoken in Ilām and adjacent areas of Iraq (i.e. MAL, BAD). Interestingly enough, Bijāri and some of the northernmost Kolyā'i dialects (e.g. Qorve) often align with the latter. The closeness of Bijāri to the vernaculars of Posht-e Kuh is indeed mentioned by Fattah (2000: 18), who states that at least a section of Bijāri speakers was resettled in its present location from Ilām sometime during the last three centuries. The whole question, I believe, merits a more detailed assessment.

A second, lesser split appears to distinguish the varieties spoken across the border with Iraq and those spoken further into Iranian territory. This distinction is identifiable in minor phonological and lexical peculiarities, consequential to the respective influence of (local and/or standard varieties of) Arabic and Persian as dominant regional languages.

The KSZ and MAL[30] dialects are the broadest subgroups but seemingly not the most internally differentiated. Their cohesion, wide diffusion and apparent high prestige might favor (or have already favored) their emergence as inter-dialectal media of spoken communication in different regions of the SK domain.[31] Observing local TV and radio broadcasts of Kermānshāh and Ilām Provinces might prove interesting in this respect.

A third distinction concerns L-KER and KOR vernaculars, standing aside from the rest of SK for their remarkable closeness to the Laki varieties of Lorestān and, especially in the case of KOR, to neighboring NLori dialects. The L-KER subgroup seems to show the highest number of divergent lexical

[29] The dialects spoken at the northern and southern peripheries diverge the most, due to contact and convergence phenomena with neighboring CK (to the north) and Laki and/or NLori dialects (to the south).

[30] Corresponding to the 'Ilāmi' and 'Malekshāhi' subgroups in Aliakbari et al. (2014: 7).

[31] Aliakbari et al. (2014: 8) mention that "the dialect of Ilam city has the greatest prestige; in encounters with speakers of Ilami, speakers of other dialects in the province tend to approximate their own speech to that of the Ilami speakers. In addition, this dialect has influenced other Kurdish dialects".

and morphosyntactic features, often aligning with Laki 'proper', as my own research has confirmed for the L-KER variety of Harsin (cf. Belelli 2016).

From this observation, a question naturally arises: are the features L-KER dialects share with other SK vernaculars sufficient to include them in the SK group, despite important similarities to Laki 'proper'? Indeed, it seems that mutual intelligibility between most L-KER dialects and neighboring Laki varieties is higher than between L-KER and other SK dialects (e.g. Kalhori of Gilān-e Gharb, personal observation). No easy answer to this classificatory challenge is currently at hand, considered the gaps in our knowledge of most SK and Laki varieties, as well as of their historical relationship.

The certain overall pattern is that, at least synchronically, the SK continuum extends south towards Laki-speaking regions of Lorestān and east Ilām, with L-KER and KOR varieties functioning as transitional links. In historical terms, we cannot tell how long this has been the case and a more in-depth study is needed before any assertion on genetic affiliations and the direction of contact induced variation in border areas can be made.

In summary, the central aim of this paper was to pose core questions that need to be addressed by future research. For a better understanding of SK dialectology, a thorough documentary activity is impelling and I agree with Rossi (1979: 162) that "more descriptive work must precede any historical one". An investigation on the extent to which native speakers perceive the classification outlined in Fattah (2000) as being grounded in reality could also prove important for making an advance in SK dialectology possible.

References

Aliakbari, Mohammad & Gheitasi, Mojtaba & Anonby, Erik. 2014. On language distribution in Ilam Province, Iran. *Iranian studies* 48(6). 835–850. (`https://doi.org/10.1080/00210862.2014.913423`).

Aliyāri Bābolqāni, Salmān. 1396/2017. *Ganjine-ye guyešhā-ye irāni: Haft guyeš az hāšiye-ye Zāgros [Collection of Iranian dialects: Seven dialects from the Zagros periphery]*. Tehran: Farhangestān-e zabān va adab-e fārsi.

Anonby, Erik. 2004–2005. Kurdish or Luri? Laki's disputed identity in the Luristan province of Iran. *Kurdische Studien* 4–5. 7–22.

Anonby, Erik & Taheri-Ardali, Mortaza, et al. (eds.). 2015–2019. *Atlas of the Languages of Iran*. Ottawa: Geomatics & Cartographic Research Centre, Carleton University. (`http://iranatlas.net/`) (Accessed 2019-06-15).

Belelli, Sara. 2016. A study on language and folklore in the city of Harsin (Kermānshāh Province, west Iran): Sketch grammar with texts and lexicon. [unpublished doctoral thesis]. "L'Orientale" University of Naples.

Blau, Joyce. 1989. Le kurde lori. *Cahier de studia iranica* 7 [Special issue: *Études irano-aryennes offertes à Gilbert Lazard*]. 37–58.

Christensen, Arthur & Barr, Kaj. 1939. *Iranische Dialektaufzeichnungen aus dem Nachlass von F. C. Andreas*. Berlin: Weidmannsche Verlagsbuchhandlung.

Darvishyān, Ali Ashraf. 1375/1996. *Farhang-e kordi-e kermānshāhi: Kordi-fārsi [Dictionary of Kermānshāhi Kurdish: Kurdish-Persian]*. Tehran: Nashr-e Sehand.

De Morgan, Jacques. 1904. *Mission scientifique en Perse: Etudes linguistiques, dialectes kurdes, langues et dialectes du nord de la Perse*. Paris: E. Leroux.

Fattah, Ismail Kamāndār. 2000. *Les dialectes kurdes méridionaux: Étude linguistique et dialectologique* (Acta Iranica 37). Leuven: Peeters.

Guizzo, Daniele. 2007. Testi nel dialetto persiano di Kermānšāh. *Annali di Ca' Foscari* 46(3). 153–174.

Izady, Mehrdad. 1992. *The Kurds: A concise handbook*. London/New York: Taylor & Francis.

Jaliliyān, Abbās. 1385/2006. *Farhang-ī bāšūr: Tāībat ba nāwčakānī kirmāšān-u īlām-u luřistān [Bashur dictionary: Specific for the regions of Kemānshāh, Ilām and Lorestān]*. Tehran: Porsemān.

Karimi-Doostan, Gholāmhossein. 1380/2001. *Kordi-e ilāmi: Barrasi-e guyeš-e badre [The Kurdish of Ilām: Study of the dialect of Badre]*. Sanandaj: University of Kurdistan.

Karimpour, Karim. 1382/2003. *Khovar Halat: Farhang-e guyeš-e kordi-e kalhori (kermānšāhi), kordi-fārsi [Dictionary of the Kalhori Kurdish (Kermānshāhi) dialect, Kurdish-Persian]*. Tāq-e Bostān: Sobh-e Rowshan.

Khan, Geoffrey. 2009. *The Jewish Neo-Aramaic dialect of Sanandaj*. Piscataway: Gorgias Press.

Khan, Geoffrey. 2011. North-Eastern Neo-Aramaic. In Weninger, S. & Khan, G. & Streck, M. P. & Watson, J. C. E. (eds.), *The Semitic languages: An international handbook*, 708–724. Berlin/Boston: De Gruyter Mouton.

Kreyenbroek, Philip Gerrit & Chamanara, Behrooz. 2013. Literary Gurāni: Koinè or continuum? In Bozarslan, H. & Scalbert-Yucel, C. (eds.), *Joice Blau: L'eternelle chez les kurdes*, 151–169. Paris: Karthala.

Lazard, Gilbert. 1992. Le dialecte laki d'Aleshtar (kurde méridional). *Studia iranica* 21(2). 215–245.

MacKenzie, David Neil. 1961a. *Kurdish dialect studies*. Vol. 1. London: Oxford University Press.

MacKenzie, David Neil. 1961b. The origins of Kurdish. *Transactions of the Philological Society* 60(1). 68–86. [repr. with *Addenda and Corrigenda* in *Iranica diversa*, Roma: IsIAO, 369–387, 674–676].

MacKenzie, David Neil. 1962. *Kurdish dialect studies*. Vol. 2. London: Oxford University Press.

MacKinnon, Colin. 2002. The dialect of Xorramābād and comparative notes on other Lor dialects. *Studia iranica* 31(1). 103–138.

Mahmoudveysi, Parvin & Bailey, Denise. 2013. *The Gorani language of Zarda, a village of west Iran: Texts, grammar, and lexicon*. Wiesbaden: Reichert.

Mahmoudveysi, Parvin & Bailey, Denise & Paul, Ludwig & Haig, Geoffrey. 2012. *The Gorani language of Gawraju (Gawrajuyi), a village of west Iran: Grammar, texts and lexicon*. Wiesbaden: Reichert.

Matras, Yaron et al. 2016. *The dialects of Kurdish*. Manchester: University of Manchester. (`http://kurdish.humanities.manchester.ac.uk/`) (Accessed 2019-05-31).

McCarus, Ernest N. 2009. Kurdish. In Windfuhr, G. (ed.), *The Iranian languages*, 587–633. London/New York: Routledge.

Morādi, Mohammad Ra'uf. 1394/2015. *Ganjine-ye guyešhā-ye irāni: Ostān-e Kermānšāh [Collection of Iranian dialects: Kermānshāh Province]*. Tehran: Farhangestān-e zabān va adab-e fārsi.

Öpengin, Ergin & Haig, Geoffrey. 2014. Regional variation in Kurmanji: A preliminary classification of dialects. *Kurdish studies* 2(2). 143–176.

Potts, Daniel. 2016. *Nomadism in Iran: From Antiquity to the Modern Era*. Oxford: Oxford University Press.

Querry, Amédée. 1896. Le dialecte Guerruci. *Mémoires de la Société de Linguistique de Paris* 9. 1–23.

Rossi, Adriano Valerio. 1979. Phonemics in Balōčī and modern dialectology. In Gnoli, G. & Rossi, A. V. (eds.), *Iranica*, 161–232. Naples: IUO.

Schmitt, Rüdiger. 2000. *Die iranischen Sprachen in Geschichte und Gegenwart*. Wiesbaden: Reichert.

5
Asymmetries in Kurmanji morphosyntax[1]

Songül Gündoğdu

Abstract: The current paper aims to investigate different morphosyntactic realization of the constituents (case vs. adposition) and their linear ordering (preverbal vs. postverbal) in a Kurmanji clause through an event structure analysis. Based on the data from Muş Kurmanji (MK), it discusses that there is a relation between the morphological form of the constituents and their status as encoded in the verb's meaning in MK; that is, structural participants are realized with case morphology while constant participants are introduced with adpositions. It further argues that the reason why MK makes a distinction in the linear ordering of structural participants is indeed a word-order property (VG) retained from proto-Kurdish and further constrained by the morphosyntactic properties of Kurmanji.

1 Introduction

Kurmanji (also known as Northern Kurdish) employs two morphological tools to indicate clausal constituents: case marking (direct vs. oblique) and adpositions (prepositions, postpositions, and circumpositions). Case marked NPs[2] generally encode event participants such as agent, patient and recipient, while adpositional phrases introduce a wide range of semantic roles

[1]A previous version of this paper was presented at at the 3rd International Conference on Kurdish Linguistics (ICKL-3), University of Amsterdam, August 25–26, 2016. I would like to thank the audiences, an anonymous referee and the volume editors for their insightful questions and detailed feedback. This project was supported by the Boğaziçi University Research fund (BAP) under Project No. 12020.

[2]Whether the case-marked noun phrases are NP (noun phrase) or DP (determiner phrase) is an issue beyond the scope of this paper. Given that such a distinction does not make any difference for the discussion here, for the sake of simplicity, I take all noun phrases as NPs.

like causee, patient recipient, benefactor, addressee, location, source and path.[3] However, case morphology and the adpositional system in this language overlap in expressing certain participant roles such as patient and recipient. Furthermore, in some instances this overlap is sensitive to the position of the constituent (preverbal vs. postverbal) in the clause. The current study investigates the following three asymmetries attested in Kurmanji clauses based on data collected in Muş – a city located in the eastern part of Turkey (henceforth, Muş Kurmanji – MK):[4]

i Verbs having an OBL patient and those having an ADP patient;[5]

ii OBL recipients with *give* and ADP recipients with *send*;

iii ADP recipients are preverbal while OBL-marked recipients and spatial goals are postverbal.

The current paper will demonstrate that the morphological realization of the constituents and their linear ordering in MK are sensitive to the correlation between verb meaning and event type. For instance, there are certain verb classes (e.g., activity verbs, motion verbs) which pattern alike with respect to argument realization properties, pointing to the existence of certain event types. Considering the MK data, I will provide an explanation for these asymmetries through an *event structure approach* which analyzes verb meaning and argument representation through event structure templates (specifically Levin & Rappaport Hovav 2007; Levin 1999, 2011; Rappaport Hovav & Levin 1998, 2000, 2008).

[3] Note that in this study, the semantic role patient is used in line with Dowty's *Patient Proto-Role*, which covers all the patient/theme properties that can function as an object in a clause. Dowty (1991: 572) specifies the contributing properties for the Patient Proto-Role as follows: "(a) undergoes change of state, (b) incremental theme, (c) causally affected by other participants, (d) stationary relative to movement of another participant and (e) does not exist independently of an event, or not at all". Furthermore, the terms *object* and *patient* are sometimes used interchangeably in this study to refer to a clausal object but none of these terms directly refer to *direct object*.

[4] According to the tentative classification of Kurmanji dialects in Öpengin & Haig (2014), Muş Kurmanji is located in the Northern Kurmanji dialect region. The MK data used in this study were collected in the form of spontaneous speech from seven native speakers of Kurmanji living in the different villages and districts of Muş during September 2015 and October 2017.

[5] Abbreviations used in this text: ADP adposition(al), DIR direct case, DIRC directional, EZ ezafe, F feminine, INDF indefinite, M masculine, NEG negation, NP noun phrase, OBL oblique case, PL plural, PROG progressive, PRS present, PST past, PTCP participial, S singular, SBJV subjunctive, V verb.

The structure of the paper is as follows: proceeding from this Introduction, Section 2 presents the asymmetries observed in MK through examples, and Section 3 proposes an event structure explanation for these asymmetries. The position of certain constituents with respect to the verb in MK will be further elucidated through a discussion on dialectal variation and contact influence in Section 4. Finally, the last section presents concluding remarks along with the issues (e.g., lexicon-syntax mapping) left for further study.

2 Asymmetries in MK morphosyntax and implications

In MK, a clausal object either carries Oblique case or it is adpositional, and this variation is sensitive to the verb type.[6] Verbs like *şikandin* 'break', *anîn* 'bring', and *xwarin* 'eat' have OBL-marked objects, while other verbs such as *hez kirin* 'like/love (lit. love do)', *temaşe kirin* 'watch', *bawer kirin* 'believe' and *nêrîn* 'look' take ADP objects. The same verb cannot mark its object with OBL or ADP in the same environment. To illustrate, the verb *şikandin* 'break' (1a) can only have an OBL object while the object of the verb *nêrîn* 'look' must be adpositional (2a). Not meeting these conditions leads to ungrammaticality, as in putative (1b) and (2b) (objects in bold).

(1) a. *Min* ***der-ê*** ***wan*** *şikand.*
1S.OBL door-EZ.M 3PL.OBL break.PST.3S
'I broke their door.'

b. * *Min* ***[ADP + derê wan]*** *şikand.*

[6] A few remarks on the case and adposition system of MK would be useful to follow the data provided in the paper easily. Just like other Kurmanji dialects, MK has a stable two-term case system: DIR and OBL. The nouns in the DIR are unmarked while those in the OBL are mostly overtly marked. Specifically, feminine singular nouns are marked by *-ê* and plural nouns by *-an* in the OBL. Although the expression of the OBL is generally absent with masculine nouns in this dialect, they have an overt OBL-marking *-î* when they are modified by a demonstrative and a quantifier, or when they are indefinite. Also note that the adpositional system of MK contains three groups of adpositional forms: basic prepositions (e.g., *ji*, *li*, *bi*, *bê*), locational nouns (e.g., *nav* 'inside', *bin* 'bottom', *ber* 'front', *ser* 'head/top', *pişt* 'behind') and postpositional particles (e.g., *ra* and *da*). Basic prepositions can be used alone as simple adpositions, such as *ji* 'from', *li* 'at', *bi* 'with' and *bê* 'without', or they can be combined with a locational noun, forming compound adpositions as in *li ber* 'in front of', *ji ber* 'because of', *li ser* 'on, upon, over', etc., or they can further be used in combination with a postpositional particle, yielding circumpositions such as *ji ... ra* 'for, to', *bi ... ra* 'together with', *di ... ra* 'through' (Gündoğdu 2018). In this paper, all types of adpositions are glossed as ADP.

(2) a. *Mêsîçî* ***li*** ***tor-ê*** *di-nêr-e.*
fisher.DIR ADP fish net-OBL PROG-look.PRS-3S
'The fisherman is looking at the fish net.'

b. * *Mêsîçî* ***torê*** *dinêre.*

Likewise, morphological marking and the position of recipients display differences based on verb type. The verbs *dan* 'give' and *firotin* 'sell' mark their recipient with OBL case and place them in the immediate postverbal position, whereas the verb *şandin* 'send' expresses its recipient through ADP in the preverbal position, as in (3)–(4) (recipients in bold face).

(3) *Ser-ê* *mal-ê* *dewar-ek-î* *bi-d-e* ***te.***
head-EZ.M house.OBL cattle-INDF-OBL SBJV-give.PRS-3S 2S.OBL
'Let each house give **you** a head of cattle.'

(4) *Xwed-ê* ***ji*** ***wî*** ***ra*** *ayet* *şand.*
God.OBL ADP 3S.OBL ADP verse send.PST.3S
'God sent **him** the verse of the Koran.'

In fact, the distinction that appears in linear ordering is not specific to the recipients of these two types of verbs. We observe a similar restriction on the distribution of other goal constituents, namely goals of verbs of movement, recipients of verbs of transfer, and addressees of verbs of speech (Haig 2014: 413). For instance, just like the recipient of *şandin* 'send', addressees also appear in preverbal position and are adpositional (5). On the other hand, goals of verbs of movement show up in postverbal position and mostly bear OBL case (6).[7]

(5) *Ez* ***ji*** ***we*** ***ra*** *meselek-î* *bi-bêj-im.*
1S.DIR ADP 2PL.OBL ADP topic.INDF-OBL SBJV-say.PRS-1S
'I will tell you about a topic.'

(6) a. *Me* *kêrî-yê* *anî* ***gund.***
1PL.OBL flock-OBL bring.PST.3S village.OBL
'We brought the flock to the village.'

[7]These verbs may sometimes take adpositional postverbal goals, although it is not common (Haig 2014). These adpositional postverbal goals are expressed by location nouns, which evolve historically from nouns (Haig 2014; Haig & Thiele 2014), like *ber* 'in front of', *nav* 'inside', *ser* 'on/above', *cem* 'next to', as in *ber min* 'in front of me', *nav malê* 'inside the home' and *cem wî* 'next to him'. Since, unlike functional prepositions, they are derived from nouns, they do not pose problems for the analysis in this study.

b. *Her sê bi hev ra ket-in* ***ç'al-ê.***
every three ADP each other ADP fall.PST-PL culvert-OBL
'All three of them fell into the culvert together.'

These examples clearly indicate that the morphological form and linear ordering of certain participant roles are totally dependent on the verb type; in other words, they reflect a distinction associated with specific verbs. The next section will focus on the lexical semantics of the verbs in order to identify the reason for this distinction.

3 An event structure proposal

Theories of argument realization (Baker 1988, 1997; Borer 1998; Cuervo 2003; Hale & Keyser 1993, 2002; Larson 1988; Levin & Rappaport Hovav 2005; Marantz 1997; Ramchand 2002, 2008, 2013, among others) aim to account for the relation between the verbs and their syntactic context by distinguishing between their structural and idiosyncratic aspects of meaning in terms of event structure and root. One intuitive idea is that verbs in sentences express events and arguments encode participants of events (Cuervo 2003). However, these theories differ in the way that arguments of a verb are projected in syntax, which aspects of verb meaning are relevant to argument realization, and how verbs get their meaning. For instance, "projectionist" approaches propose that argument structure of a verb is projected into syntax through theta-role assignment and subcategorizational features (Baker 1988, 1997; Bresnan 1982; Chomsky 1981; Grimshaw 1990; Larson 1988). The idea at the heart of this view is that there is a lexicon where each verb is stored with semantically (e.g., theta roles) and syntactically (e.g., number of arguments) relevant information, and that the argument structure of a verb is determined based on this lexical information. On the other hand, "constructivist" approaches take the opposite view of argument structure, emphasizing the idea that verb meaning resides in the syntactic context. That is, the lexical entry of a verb registers only its core meaning (root) and the meaning of a verb is determined compositionally within the syntactic structure it builds up (Borer 2005; Chomsky 1995; Hale & Keyser 2002; Halle & Marantz 1993; Marantz 2013; Ramchand 2008). Although they seem different, the main idea of both approaches is similar: each verb has its own argument structure realization, either stored in the lexicon or determined within the syntactic context.

The MK data clearly demonstrate that certain groups of verbs pattern alike with respect to argument realization properties. This implies that there are a number of verb classes which share the same semantic structure which in turn determines their morphosyntactic realization. Therefore, an event structure approach that takes a number of primitive predicates (e.g., ACT, CAUSE, BECOME, etc.) to determine the event type of certain verb classes, and their grammatical behaviors seems superior to argument realization approaches that treat every verb differently.[8] In order to account for the morphosyntactic asymmetries observed in this dialect, I will draw on the event-structure-based approaches in the literature, specifically from those proposed by Levin (1999, 2011), Levin & Rappaport Hovav (2007), and Rappaport Hovav & Levin (1998, 2008).

In their work, Rappaport Hovav & Levin (1998) argue that event structure denotes the representation of verb meaning and determines various grammatical properties, including the realization of arguments. In their approach, the meaning of a verb is bipartite: event structure and core meaning. The former refers to the structure that the verbs share with other verbs of the same semantic type, so it is the *structural facet* of verb meaning which defines the possible event types. In contrast, the latter is directly relevant to what is idiosyncratic to that verb, thus it is the *idiosyncratic facet* of verb meaning that differentiates one verb from others sharing the same structural facets of meaning (i.e., constant).[9] The authors assume a small set of event structure templates that contain the inventory of possible event types, which are to some extent aspectually motivated, namely simplex and complex event structure templates. Simplex event templates consist of one single sub-event whereas complex event templates contain two sub-events:

(7) Simple event structure templates

a.	[x ACT<*MANNER*>]	(activity)
b.	[x <*STATE*>]	(state)
c.	[BECOME [x <*STATE*>]]	(achievement)

[8] Note that some constructivist approaches integrate the event structure templates into syntax successfully through an "event decompositional" syntactic model. See Cuervo (2003), Hale & Keyser (1993, 2002), and Harley (1995) for sample applications and further discussion.

[9] In fact, such a distinction also exists in other verb-meaning-based argument structure approaches; for instance, *structural facet* corresponds to *semantic structure* of Grimshaw (1990), or *structural configuration* of Hale & Keyser (1993). Likewise, *idiosyncratic facet* is the *semantic content* or *head* inserted in the structure in these studies, respectively.

(8) Complex event structure template
[[x ACT<*MANNER*>] CAUSE [BECOME [y <*STATE*>]]] (causative)
(Levin 1999: 9)

It is crucial to specify that two types of participants are encoded in an event structure: "structural participants", which are required as well as licensed by virtue of both the event structure template and by the verb meaning, and "constant participants", which are only required and licensed by virtue of the constant alone.[10] In Levin's work, structural participants are expressed by variables as "x" and "y" and constants are indicated as underlined variables such as "y". The main idea is that simplex event templates have only one structural participant but may have one or more constant participants based on the idiosyncratic meaning of the verb. Complex event templates have two structural participants and may have constant participants if licensed. For instance, *sweep* is an activity verb that needs minimally a *sweeper* and a *surface*, hence its meaning is associated with two participants: the structural participant *sweeper* and the constant participant *floor*, as in *I swept the floor*. Similarly, a causative (or accomplishment) verb such as *break* has two structural participants: the actor who breaks and the undergoer which is broken.

3.1 Two types of verbs

The fact that verbs introduce their objects in different morphological forms is not specific to MK or Kurmanji in general. Croft (1993) points out that although languages are not uniform in argument realization of non-causative psych-verbs (e.g., *fear*), they are consistently uniform in the argument expression of causative psych-verbs (e.g., *frighten*). Levin (1999) also observes that languages are uniform in expressing the arguments of causative verbs such as *cut*, *kill* and *break*, but they display variation in the argument realization of non-causative verbs in general like *sweep*, *greet* and *answer*. The object(-like) arguments of these latter verbs show more than one potential morphosyntactic realization in English and across languages. Levin proposes that verbs with complex/causative event structures are core transitive verbs (CTV): they are obligatorily transitive, since they have two structural participants required by the event structure template, and these participants are

[10] Grimshaw & Vikner (1993) also establish a dichotomy between arguments based on their behavior: structure arguments are licensed by semantic structure while content arguments are licensed by the semantic content.

mapped onto syntax as subject and direct object. On the other hand, non-causative verbs are two-argument verbs with simplex event structures. They are non-core transitive verbs (NCTV) thus they may – but need not – be transitive as the constant participant (i.e., the argument licensed by the verb's core meaning) does not fall under the event structure-to-syntax mapping principle and is generally realized as oblique argument in syntax. In fact, the contrast that we observe in the morphosyntactic realizations of objects in MK is similar to the distinction between CTV and NCTV made by Levin (1999). This contrast stems from the fact that the objects in this dialect differ in their status with respect to their source in the event structure template.

When we look at verbs with ADP objects in MK, we see that they are all activity verbs like *nêrîn* 'look' (simplex verbs) and *temaşe kirin* 'watch', *hez kirin* 'like/love (lit. love do)', *se'h kirin* 'listen' (complex verbs). The significant point is that the objects of these verbs do not carry the properties of a typical direct object in Kurmanji. Direct objects in this language are non-adpositional and achieve subjecthood under passivization (Haig 2002: 20) as illustrated in (9). On the contrary, ADP objects are always adpositional (2a is repeated as 10a) and they cannot be the subject of the passivized verb (10b).

(9) a. *Zarok-an* ***pişîk-ê*** *kuşt.* (ACTIVE)
child-PL.OBL cat-OBL kill.PST.3S
'The children killed the cat.'

b. ***Pişîk*** *hat* *kuştin.* (PASSIVE)
cat.DIR come.PST.3S kill
'The cat was killed.'

(10) a. *Mêsîçî* ***li*** ***tor-ê*** *di-nêr-e.*
fisherman.DIR ADP fish net-OBL PROG-look.PRS-3S
'The fisher is looking at the fish net.'

b. * ***li*** ***tor-ê*** *hat* *nêrîn*
ADP fish net-OBL come.PST.3S look

However, verbs with ADP objects behave parallel to the verbs with true direct objects with respect to the ergative alignment in past tense constructions.[11] Based on this observation, Haig (2002) makes a distinction between

[11] Kurmanji displays an ergative pattern in past tense constructions, where the subject of an intransitive verb (S) is treated similarly to the object of a transitive verb (O) and differently from the transitive subject (A); thus, transitivity and intransitivity of the verb in this language are assessed with respect to ergative alignment in past tense constructions.

clausal and lexical transitivity in Kurmanji, proposing that "only transitive verbs can govern a direct object; intransitive verbs cannot. However, not all transitive verbs govern a direct object" (2002: 20). According to this classification, transitive verbs have direct objects while lexically transitive verbs do not; but the latter group licenses ADP objects. I argue that in MK, verbs with ADP objects are indeed non-core transitive verbs; they are all single activity verbs with simple event templates consisting of two participants: structural and constant participants.[12] The structural participant of these verbs is the doer of the action (actor or initiator) and they are morphologically realized as a case-marked NP. The constant participant of these verbs, on the other hand, may be a person, a thing, a location or manner (oblique argument) and their morphological realization is an ADP phrase.[13] For instance, the event template of a NCTV like *nêrîn* 'look' can be expressed as follows: (Note that 'y' stands for the constant participant in (11).)

(11) a. [x ACT<*MANNER*> y]
b. [x ACT<*NÊRÎN*> y]
c. [SUBJECT ACT<*NÊRÎN*> ADP OBJECT]

Core transitive verbs with complex event structure templates have obligatory OBL objects, which qualify as true direct objects in MK. They have two structural participants: actor (subject) and undergoer (direct object), both of which are morphologically realized as case-marked NPs. The event template of a verb like *şikandin* 'break' would be a good example of a CTV. It should be noted that what is idiosyncratic to a CTV is the state it lexicalizes, thus *şikestî* 'broken' in (12b) is the state that the event *şikandin* 'break' lexicalizes in its event structure template. (12) roughly means that there is an external causer (i.e., subject) which acts upon an object (i.e., undergoer) and changes its state.

(12) a. [[x ACT<*MANNER*>] CAUSE [BECOME [y *STATE*]]] (causative)
b. [[x ACT <*MANNER*>] CAUSE [BECOME [y *ŞİKESTÎ*]]]
[SUBJECT ACT<*MANNER*> CAUSE [BECOME OBL OBJECT *ŞİKESTÎ*]]

[12] Note that stative verbs like *zanîn* 'know' in MK also have ADP objects, which is in line with the fact that stative verbs represent simple events just like single activity verbs.

[13] The majority of non-core transitive verbs in this dialect constitute complex predicates which are classified as "unergative complex predicates" in Gündoğdu (2016). Due to space limitation I will not elaborate on them here but the reader is referred to this study for a detailed syntactic account of these verbs.

However, there are a few verbs such as *dîtin* 'see' and *xwandin* 'read' which are definitely not core transitive verbs but nevertheless license OBL objects in MK. In fact, these verbs behave like a core transitive verb in terms of morphological marking of their objects in many languages (e.g., English, Turkish, Persian, Japanese, Basque, Warrungu, etc.) (Tsunoda 1985). It seems that such verbs have a strong preference for a transitive syntactic frame in these languages, and this is why they require their object to be Oblique-marked just like core transitive verbs in MK (as throughout Kurmanji).

3.2 Two types of recipients

As stated in Section 2, the form of the recipients in this dialect is also sensitive to the verb type, as *give*-type verbs have OBL recipients while *send*-type verbs have ADP recipients. I will argue that this difference is due to the fact that these verbs lexicalize different properties of "transfer" information in their event structure, hence this distinction is morphosyntactically reflected.

Investigating the different argument realizations of three-participant constructions such as *give*, *sell*, *send* in dative alternations across languages under the "verb sensitivity approach", Levin & Rappaport Hovav (2007) and Levin (2011) argue that *give*-type verbs (e.g., *give*, *sell*, *hand*, *rent*) inherently lexicalize only caused possession in their meaning. Therefore, these verbs are only associated with the change of possession or "caused possession" event type (13). On the other hand, *send*-type verbs (e.g., *send*, *mail*, *ship*, etc.) inherently lexicalize spatial goals and thus their roots are associated with caused motion as they denote a physical change of location of the theme (14):

(13) Caused possession:
'X$_{\text{AGENT}}$ ACT CAUSE Y$_{\text{RECIPIENT}}$ HAVE Z$_{\text{THEME}}$'
(adapted from Levin & Rappaport Hovav 2007)

(14) Caused motion:
'X$_{\text{AGENT}}$ ACT CAUSE Z$_{\text{THEME}}$ BE LOC Y$_{\text{SPATIAL GOAL}}$'
(adapted from Levin 2011)

Both event templates have three inherently involved participants, but they differ in lexicalizing the participant that denotes a change: *give*-type verbs lexicalize the agent, theme and recipient (change of possession), whereas *send*-type verbs lexicalize the agent, theme and spatial goal (change of location). In English, the participant roles in the caused possession event type

can give rise to two syntactic configurations, namely (a) double-object construction (DOC) – V NP NP, and (b) the *to*-prepositional ditransitive variant – V NP to NP.

(15) a. *Sandy gave Terry a copy of the new grammar.*
b. *Sandy gave a copy of the new grammar to Terry.*
(Levin & Rappaport Hovav 2007: 1)

Nevertheless, this event type lacks a conceptual path and thus it does not entail a physical transfer of possession from a source to a goal/recipient but rather it merely denotes a change of possession taking place between the original possession and the recipient. Therefore, in both syntactic configurations, only the caused possession is encoded, regardless of the recipient being realized as the first object in DOC or as the complement of the preposition *to* (Levin & Rappaport Hovav 2007; Levin 2011). Even though the roots of *send*-type verbs do not inherently lexicalize caused possession, they may be associated with the caused possession in some languages, e.g., English:

(16) a. *Mary sent some newspapers to the library.*
(spatial goal-caused motion)
b. *Mary sent some newspapers to Jane.*
(caused motion or caused possession where *Jane* is interpreted as a recipient)
c. *Mary sent Jane some newspapers.*
(caused possession where *Jane* is interpreted as the recipient)

The basic distinction between the event structures of these two verb types, based on which participant role is lexicalized, is crucial for the MK data. Morphological marking seems to point to a distinction between structural and constant participants in MK: structural participants are realized with case morphology while constant participants are expressed by adpositions. If this is the case, then we expect to find that the recipient of *dan* 'give' (and also *firotin* 'sell') appears in OBL as it lexicalizes caused possession, whereas the recipient of *şandin* 'send' is expressed through ADP since it does not lexicalize caused possession. Furthermore, since *şandin* 'send' lexicalizes *caused motion* as it refers to a physical change of location, we expect to find the location as an OBL spatial goal – the constituent denoting the spatial endpoint of the

event. This is what we get in MK; *give*-type verbs have OBL recipients (17) while *send*-type verbs have ADP recipients but OBL spatial goals (18):[14]

(17) *ser-ê* *mal-ê* *dewar-ek-î* *bi-d-e* ***te.***
head-EZ.M house.OBL cattle-INDF-OBL SBJV-give.PRS-3S 2S.OBL
'Let each [person] give **you** a head of cattle.'

(18) *Min* *nan-ê* *wan* ***ji*** ***wan*** ***ra*** *şand* ***zevî-yê.***
1S.OBL bread-EZ.M 3PL.OBL ADP 3PL.OBL ADP send.PST.3S field-OBL
'I sent them their meal to the field.'

To summarize, the fact that the recipients of *give*-type verbs and of *send*-type verbs carry different morphology is not arbitrary, but rather is sensitive to the event structure of these verbs; more specifically, it depends on whether the recipient is the structural participant (i.e., inherently lexicalized) or the constant participant (i.e., licensed by the idiosyncratic meaning of the verb). However, the reason why the recipients of these verbs appear in different positions within the clause still needs explanation.

3.3 Two types of positions for goal constituents

Levin (2011) observes that the actual realization of the caused possession and caused motion event schema shows differences across languages due to different types of morphosyntactic resources that languages make use of for expressing these schemata. She finds that (i) some languages have the same realization for both goals and recipients while (ii) in other languages there are two realizations for recipients, one of which is shared by the goal and (iii) still other languages allow two realizations of goals, one of which is the same as the recipient. As illustrated in (17) and (18), the event schema of three participant verbs such as *give* and *send* in MK corresponds to the morphosyntactic realization attested in type (ii) languages (19):

[14] In fact, the verb *dan* 'give' can also lexicalize a caused motion event structure in MK, because we find sentences where *dan* 'give' has a spatial goal, which is OBL-marked occurring in the postverbal position:

Kimik-ê *bi-d-e* *ser-ê* *xwe!*
cap-OBL SBJV-give.PRS-2S head-EZ.M. self
'Wear the cap!' (lit.: 'Give the cap on your head!)'

However, in such cases, *dan* 'give' semantically patterns with the verb *danîn* 'put' rather than expressing a giving event. See Gündoğdu (2018) for further discussion.

(19) a. *dan* 'give' = NP$_{\text{AGENT}}$ NP$_{\text{OBJECT}}$ VERB$_{\text{GIVE}}$ **NP$_{\text{RECIPIENT}}$**

b. *şandin* 'send' = NP$_{\text{AGENT}}$ **PP$_{\text{RECIPIENT}}$** NP$_{\text{OBJECT}}$ VERB$_{\text{SEND}}$ **NP$_{\text{SPATIAL GOAL}}$**

The recipient of *dan* 'give' and the spatial goal of *şandin* 'send' appear in the same position and as the structural participants of these verbs; they are both OBL-marked. However, the recipient of *şandin* 'send' shows up in the preverbal position and is introduced with an ADP as the constant participant of this verb.

So far, what we observe in MK, viewed through the lens of the proposals of the event structure approach (along with the claims of the verb sensitivity approach), is that this dialect reflects the distinction between structural and constant participants not only through morphology (case vs. adposition) but also through the position of the constituent with respect to the verb (pre- vs. post-predicate). Therefore, structural participants other than the actor/initiator and patient/theme/undergoer appear in the immediate postverbal position of the clause, e.g., the recipient of *dan* 'give' and the spatial goal of *şandin* 'send'; whereas the constant participants appear in the preverbal position. This proposal receives further support from other goal constituents in MK. Recall from Section 2 that the addressee patterns alike with the recipient of *şandin* 'send'; both are adpositional and appear in the preverbal position. Similarly, the goals of verbs of movement display the same properties as the spatial goal of *şandin* 'send': they bear Oblique case and show up in the immediate postverbal position. Therefore, the MK data points to two different goal positions in the sentence (20):

(20) **G**OAL (RECIPIENTS, ADDRESSEE) **V**ERB **G**OAL (RECIPIENTS, SPATIAL GOAL)

To this end, I wish to address the following question: How should we approach addressee constituents in the preverbal position and spatial goal constituents in the postverbal position? I suggest that the addressees of speech verbs showing up in the preverbal goal position in MK are not inherently lexicalized, and thus they are constant participants just like the recipient of *şandin* 'send'. In other words, the addressee is not, in fact, a part of the event structure; whereas the meaning of the verb *gotin* 'say/tell' already implies the presence of a hearer or listener.

Likewise, in addition to *şandin* 'send', there is a group of verbs of movement that place their spatial goals in the postverbal position. Below is a list of these verbs attested in the MK data:

(21) Verbs with spatial goals in MK:

a. *anîn* 'bring'
avetin 'throw'
birin 'carry, take to somewhere'
danîn 'put, leave'
xistin 'put'
berdan 'release'

b. *derketin* 'go out, leave'
hatin 'come, arrive'
ketin 'fall, enter'
çûn 'go'

The verbs in (21a) are just like the verb *şandin* 'send' in terms of their event schema because each denotes a physical change of location of the object as a result of a caused motion. Therefore, their spatial goals are indeed inherently lexicalized and licensed both by the verb root and its event structure. On the other hand, the verbs in (21b) are all path verbs (Levin & Rappaport Hovav 1995; Kudrnáčová 2008) which obligatorily encode the directionality of the motion, thus Rosen considers this type of verbs to be "verbs of inherently directed motion" (1984: 74). In fact, the inherent directionality of path verbs in general necessitates a spatial grounding or an achieved location (the second type of result verbs in Rappaport Hovav & Levin 1998). Kudrnáčová (2008: 35) explains the semantics of path verbs as follows: "they express pure translation by specifying the motion of an entity as changes in the entity's positions with respect to a spatial reference point". Consequently, the directional path encoded in path verbs is obligatory and non-additive. This suggests that the spatial goals of path verbs are also licensed by the verb root as well as its event structure. In fact, all spatial goals in MK carry the properties of a final state of the event or the result subevent (resultee), in terms of Ramchand (2002, 2008), which means that they are not only encoded in the event schemata but are also expressed as an argument of the predicate in syntax. Based on this observation, I take spatial goals of path verbs as the same as the spatial goals of verbs of caused motion in MK and claim that they are all linked to a position in the event structure of the verb as its structural participants. This explains why they usually carry case morphology in the same way as the structural participants do in MK and similarly why they occur

in the immediate postverbal position just like other non-actor/non-patient structural participants in this dialect.

4 Further issues: Dialectal variation and language contact

The morphosyntactic asymmetries attested in MK demonstrate that the distinction between structural vs. constant participants is reflected morphologically (case vs. adposition) and linearly (preverbal vs. postverbal):

Structural participants (OBL) Constant participants (ADP)	VERB	Structural participants (OBL)

However, the question as to why some structural participants appear in the preverbal position while others are placed in the postverbal position is still unresolved. Given that structural participants are already distinguished by case morphology from other types of participants, why does MK need to make a further distinction between structural participants through linear ordering in a clause? The phenomenon of postverbal goals and variation observed in their positions in other Kurmanji dialects provide us with insights that help to answer to this question.

The morphological coding and linear positioning of participants display variation across Kurmanji dialects.[15] This variation is mostly conditioned by language contact (Haig 2014) and areal linguistic typology (Stilo 2005; 2009), and the distribution of goals in all dialects is sensitive to the verb type. However, in addition to the verb type, the morphological form of the goals as well as the type of the adpositions that goals are expressed by (preposition vs. circumposition vs. postposition) seem to have an influence on this distribution. To the best of my knowledge, spatial goals of verbs of motion are always in the immediate postverbal position in all Kurmanji dialects, since an allative reading is available only in this position;[16] however, their morphological form may vary depending on the dialect region. Some dialects such as Muş and Malatya prefer OBL spatial goals while other dialects like Hakkari and

[15] See Haig (2014) and Haig & Thiele (2014) for more examples and extensive discussion on the regional variation observed in Kurdish with respect to preverbal and postverbal goals.

[16] "Allative" expresses a motion to or toward a given referent.

Şırnak tolerate both ADP and OBL spatial goals.[17] Furthermore, in all dialects, the recipient of *give*-type verbs is almost always OBL-marked and appears in the immediate postverbal position[18], while the recipient of *send*-type verbs is always adpositional and shows up in the preverbal position. In contrast, the morphological form and the position of addressee display variation. For instance, unlike MK ((5) is repeated as (22)), in the southeastern section of Kurmanji (in and around Duhok and Hakkari provinces), addressees are case-marked and postverbal (23):

(22) *Ez* ***ji*** ***we*** ***ra*** *meselek-î* *bi-bêj-im.*
1S.DIR ADP 2PL.OBL ADP topic.INDF-OBL SBJV-say.PRS-1S
'I will tell you about a topic.'

(23) *Henê* *meselek* *gût-e* ***min.***
Henê.OBL topic.INDF.DIR say.PST.3S-DIRC 1S.OBL
'Henê told me about a topic.'

Drawing attention to the fact that the appearance of goal constituents (G) in the immediate postverbal position in an OV language like Kurdish is typologically unusual, Haig (2014) and Haig & Thiele (2014) assert that this unusual word order (OVG) emerges as a result of contact-induced change. Haig (2014) argues that an original 'proto-Kurdish' had V(erb)G(oal) order which was characterized through early Aramaic/Iranian contact. In due course this pattern has undergone changes in some Kurmanji dialects due to contact with various languages. For instance, in the southernmost Kurmanji dialects, VG order has been mostly preserved due to the contact with Neo-Aramaic, which is a VO language, and thus goals are predominantly postverbal. On the other hand, goals are overwhelmingly preverbal in the Kurmanji dialects to the north and west (which Haig labels as Central Anatolian dialects) because of the influence of Armenian and Turkish varieties, both of which are OV languages. In keeping with his analysis, MK is one of the dialects in which

[17] The data for this section come from 13 Kurmanji speakers living in different districts of Hakkari, Van, Şırnak, Mardin, Muş, Bingöl, Malatya and Adıyaman provinces. 11 of them were undergraduate students at Muş Alparslan University and 2 of them were working in Malatya when the data were elicited.

[18] Malatya Kurmanji is exceptional to some extent because not all recipients of all *give*-type verbs exhibit the same properties; e.g., the recipient of *dan* 'give' is adpositional and appears in the preverbal position, whereas the recipient of *firotin* 'sell' is OBL and postverbal. I will mention this distinction while discussing the example given in (24).

certain goals have been shifted to preverbal position due to language contact with Armenian and Turkish whereas southeastern dialects like Hakkari Kurmanji mostly preserve the VG order retained from proto-Kurdish.

Furthermore, variation across dialects also has something to do with the adpositional system of a given dialect. Despite being an OV language where postpositions are the norm (Dryer 2013), Kurmanji has prepositions, postpositions and circumpositions. Stilo (2005, 2009) proposes that Iranian languages are sandwiched between prepositional (Semitic) and postpositional (Turkic, Armenian, Indic) patterns, and they resolve this conflict by creating an intersection zone which accommodates both patterns. As an Iranian language bordering an area between prepositional Neo-Aramaic and postpositional Armenian and Turkic, Kurmanji has both opposite typologies (e.g., preposition and postposition) and a hybridized pattern formed by the merge of these two opposites (e.g., circumpositions). The southernmost dialects (e.g., Hakkari) are mostly prepositional and goals are predominantly postverbal; on the other hand, northern and western dialects have circumpositions or independent postpositions and use both pre- and postverbal positions actively to disambiguate goal types. In fact, the dialects that shift certain goal constituents to the preverbal position as a result of language contact (Haig 2014) are those that have developed circumposition *ji … ra* (e.g., Muş) or independent postposition *… ra* (e.g., Malatya). For instance, MK introduces addressee (22) through *ji … ra*, while Malatya Kurmanji places both the addressee and the recipient of the verb *dan* 'give' in the preverbal position (24).[19] Note that the goal arguments introduced within circumpositions are always preverbal, and postverbal placement of such phrases is not an option.

(24) a. *Bahar-ê vaha **mi** **ra** ne-got.*
Bahar-OBL as such 1S.OBL ADP NEG-say.PST.3S
'Bahar didn't tell **me** like that.'

[19] There is also an example from Şahînê Bekirê Soreklî's book *Roja dawîn ji jiyana Mistê kurê Salha Temo* (1982) in which the recipient of the verb *dan* 'give' is introduced with the preposition *bi* 'with' in the preverbal position. The author of the book is from Kobanê, Syria. (I would like to thank Ergin Öpengin for bringing this sentence to my attention.)

> *Min sandiqeke tijî şûşeyên kazozê li dera hanê dî. Were ez yekê **bi te** dim; bi rê va vexwe…*
> ('I saw there a box full of bottles of soft drink. Come, I **give you** one; drink it on your way…')

b. *Zana* ***Rojday*** ***ra*** *kitap da-y-e.*
Zana Rojda.OBL ADP book give.PST-PTCP-COP.3S
'Zana gave the book **to Rojda**.'

As mentioned previously, the recipient of *send*-type verbs is preverbal in all dialects. Assuming that the addressee is also a type of recipient (Goldberg 1995),[20] northern and western dialects demonstrate that there is a tendency to shift human goals expressed by recipient roles to the preverbal position while reserving the postverbal position for locational/spatial goals for allative reading.

The data provided here demonstrate that OBL-marked goal constituents do not survive in the preverbal domain in all Kurmanji dialects, and that those appearing in the preverbal position are always adpositional. This observation implies that the linear order of goal constituents in this language is sensitive to morphological marking. It seems that more than two OBL-marked constituents cannot be licensed in the preverbal domain; therefore, a third case-marked constituent is obligatorily placed in the postverbal position. I propose that Kurmanji imposes the following general restriction on the linear order of constituents:

(25) *In a Kurmanji clause, at most two case-marked NPs (subject and direct object) are licensed in the preverbal position.*

To sum up, dialectal variation suggests that goal constituents appear in the postverbal position in Kurmanji due to the VG order retained from proto-Kurdish. The reason why some goal constituents are shifted to the preverbal position in some dialects is because of the contact these dialects have had with OV/postpositional languages such as Armenian and Turkic. MK is one of the dialects which has a long history of language contact with Armenian and Turkish, and has thus developed circumpositions, and has shifted addressees to the preverbal position. I suggest that, as a result of language contact, MK has adapted its synchronic grammar in such a way that it ends up with a morphological and linear distinction between structural and constant participants.

[20] In many languages, addressees of speech verbs are marked with dative case or with dative-like prepositions. This overlap is generally explained through addressee being construed as the recipient of the speech act, which is indeed the information being transferred.

5 Conclusion

In this paper, I have provided an event structure analysis in order to account for salient asymmetries attested in Kurmanji morphosyntax. The data have demonstrated that there is a relation between the morphological form of the constituents and their status as encoded in the verb's meaning in MK; that is, structural participants are realized with case morphology while constant participants are introduced with adpositions. Therefore, I have proposed that the objects of core transitive verbs are structural participants and thus are marked with OBL case. In contrast, the objects of non-core transitive verbs (or single activity verbs) are constant participants and thus are expressed through adpositions. Similarly, although both *send*-type verbs and *give*-type verbs license recipients as their event participants, the status of the recipient is different in the event structures of these verb groups. Recipients of the former group (caused-motion verbs) are not inherently lexicalized but are only licensed by the verb meaning, hence it is a constant participant. The recipients of the verbs in the latter group (caused-possession verbs) are inherently lexicalized thus are the structural participants of the verbs. This distinction is also morphologically reflected in MK; the recipient of *send* is adpositional, while the recipient of *give* is case-marked. Further evidence for the distinction between structural and constant participants comes from other participant roles (i.e., goal constituents); that is, the spatial goals of motion verbs and path verbs are structural participants and they are case-marked whereas the addressees of speech verbs are expressed with adpositions, since they are constant participants.

Furthermore, I have suggested that the reason why MK makes a distinction in the linear ordering of structural participants is indeed a word-order property (VG) retained from proto-Kurdish and further constrained by the morphological properties of Kurmanji. The data from other dialects as well as the findings in the literature have demonstrated that postverbal goals are preserved in the southernmost dialects, while certain goal constituents have been shifted to the preverbal position in certain dialects under the influence of contact with OV languages. I have argued that MK is one of these dialects and that it adapts its synchronic grammar in such a way that it ends up with a distinction between structural and constant participants – a distinction that is reflected morphologically and linearly.

The current paper does not deal with the event structure–syntax mapping, however. Given that syntactic structure is built on the information (event type, number of participants, etc.) encoded in a verb's event structure, the

structural vs. constant participant distinction must be preserved within the syntactic configuration. Rappaport Hovav & Levin (1998: 113) propose the following principle to regulate event structure–syntax mapping:

(26) The Structure Participant Condition: There must be an argument XP in the syntax for each structure participant in the event structure.

Thus, each structural participant must be expressed in syntax, whereas constant participants may but need not be expressed in the syntactic structure without violating this principle. Constant participants may have their syntactic expression through language-specific rules (oblique–argument linking rule). The fact that structural participants are case-marked in MK while constant participants are expressed by adpositions implies that the status of the event participants is preserved within the syntactic structure. It can be asserted at this point that structural participants check their case feature at relevant functional heads while constant participants as oblique arguments get their case feature checked by the adposition head. However, the position where event participants merge in syntax and their possible case relations within this configuration are issues that necessitate further study.

References

Baker, Mark. 1988. *Incorporation: A theory of grammatical function change*. Chicago: University of Chicago Press.

Baker, Mark. 1997. Thematic roles and syntactic structure. In Haegeman, L. (ed.), *Elements of grammar: Handbook in Generative Syntax*, 73–137. Dordrecht: Kluwer.

Borer, Hagit. 1998. Deriving passive without theta roles. In Lapointe, S. G. & Brentari, D. K. & Farrell, P. M. (eds.), *Morphology and its relation to phonology and syntax*, 60–99. Stanford, CA: CSLI.

Borer, Hagit. 2005. *Structuring sense*. Vol. 2: *The normal course of events*. New York: Oxford University Press.

Bresnan, Joan. 1982. *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa lectures*. Dordrecht: Foris.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Croft, William. 1993. Case marking and the semantics of mental verbs. In Pustejovsky, J. (ed.), *Semantics and the lexicon* (Studies in Linguistics and Philosophy 49), 55–72. Dordrecht: Springer Netherlands. (https://doi.org/10.1007/978-94-011-1972-6_5).

Cuervo, María Cristina. 2003. *Datives at large*. Massachusetts Institute of Technology dissertation. (http://hdl.handle.net/1721.1/7991) (Accessed 2018-12-04).

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Dryer, Matthew S. 2013. Relationship between the order of object and verb and the order of adposition and noun phrase. In Dryer, M. S. & Haspelmath, M. (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/95) (Accessed 2019-03-27).

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. University of Chicago Press.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Grimshaw, Jane & Vikner, Sten. 1993. Obligatory adjuncts and the structure of events. In Reuland, E. & Abraham, W. (eds.), *Knowledge and language*, 143–155. Dordrecht: Springer Netherlands. (https://doi.org/10.1007/978-94-011-1842-2_7).

Gündoğdu, Songül. 2016. Noun-verb complex predicates in Kurmanji Kurdish: A syntactic account. In Bellamy, K. & Karvovskaya, E. & Kohlberger, M. & Saad, G. (eds.), *Proceedings of ConSOLE XXIII*, 7–9 January 2015, Paris, 383–405. Leiden: Leiden University Centre for Linguistics. (https://www.universiteitleiden.nl/binaries/content/assets/geesteswetenschappen/lucl/sole/console23-updated.pdf) (Accessed 2019-08-28).

Gündoğdu, Songül. 2018. *Argument-adjunct distinction in Kurmanji Kurdish*. Boğaziçi University, Istanbul dissertation.

Haig, Geoffrey. 2002. Noun-plus-verb complex predicates in Kurmanji Kurdish: Argument sharing, argument incorporation, or what? *Sprachtypologie und Universalienforschung (STUF)* 55(1). 15–48.

Haig, Geoffrey. 2014. VG word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In Khan, G. & Napiorkowska, L. (eds.), *Neo-Aramaic and its linguistic context*, 407–425. New York: Gorgias Press.

Haig, Geoffrey & Thiele, Hanna. 2014. Post-predicate goals in Northern Kurdish and neighboring languages: A pilot study in quantitative areal linguistics. Paper presented at the *Second International Conference on Variation*

and Change in Kurdish (VCK-2), 08–09 October 2014, Mardin Artuklu University, Turkey.

Hale, Kenneth & Keyser, Samuel Jay. 1993. On argument structure and the lexical expression of syntactic relations. In Hale, K. & Keyser, S. J. (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 53–109. Cambridge, MA: MIT Press.

Hale, Kenneth & Keyser, Samuel Jay. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.

Halle, Morris & Marantz, Alec. 1993. Distributed morphology and the pieces of inflection. In Hale, K. & Keyser, S. J. (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.

Harley, Heidi B. 1995. *Subjects, events, and licensing*. Massachusetts Institute of Technology dissertation.

Kudrnáčová, Naděžda. 2008. *Directed motion at the syntax-semantics interface*. Brno: Masaryk University.

Larson, Richard K. 1988. On the double object construction. *Linguistic inquiry* 19(3). 335–391.

Levin, Beth. 1999. Objecthood: An event structure perspective. In Billings, S. & Boyle, J. & Griffith, A. (eds.), *Proceedings of CLS 35*, vol. 1: *The main session*, 223–247. Chicago: Chicago Linguistic Society, University of Chicago.

Levin, Beth. 2011. Verb sensitivity and argument realization in three-participant constructions: A crosslinguistic perspective [conference handout]. *Conference on Referential Hierarchies in Three-participant Constructions*, 20–21 May 2011, Lancaster University. (`http://web.stanford.edu/~bclevin/lanc11dat.pdf`) (Accessed 2019-01-03).

Levin, Beth & Rappaport Hovav, Malka. 1995. *Unaccusativity*. Cambridge, MA: MIT Press.

Levin, Beth & Rappaport Hovav, Malka. 2005. *Argument realization*. New York: Cambridge University Press.

Levin, Beth & Rappaport Hovav, Malka. 2007. The crosslinguistic study of dative alternations: A verb sensitive perspective [conference handout]. *Conference on Ditransitive Constructions*, 23–25 November 2007, Max Planck Institute for Evolutionary Anthropology, Leipzig.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics* 4(2). 201–225.

Marantz, Alec. 2013. Verbal argument structure: Events and participants. *Lingua* 130. 152–168.

Öpengin, Ergin & Haig, Geoffrey. 2014. Regional variation in Kurmanji: A preliminary classification of dialects. *Kurdish studies* 2(2). 143–176.
Ramchand, Gillian. 2002. Aktionsart, L-syntax and selection. In Verkuyl, H. J. (ed.), *Proceedings of Perspectives on Aspect Conference*, 1–15. Utrecht: OTS.
Ramchand, Gillian. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge, UK: Cambridge University Press.
Ramchand, Gillian. 2013. Argument structure and argument structure alternations. In den Dikken, Marcel (ed.), *The Cambridge handbook of Generative Syntax*, 265–321. Cambridge, UK: Cambridge University Press.
Rappaport Hovav, Malka & Levin, Beth. 1998. Building verb meanings. In Butt, M. & Geuder, W. (eds.), *The projection of arguments*, 97–134. Stanford, CA: CSLI Publications.
Rappaport Hovav, Malka & Levin, Beth. 2000. Classifying single argument verbs. In Coopmans, P. & Everaert, M. & Grimshaw, J. (eds.), *Lexical specification and insertion*, 269–304. Amsterdam: Benjamins.
Rappaport Hovav, Malka & Levin, Beth. 2008. The English dative alternation: The case for verb sensitivity. *Journal of linguistics* 44(1). 129–167.
Rosen, Carol. 1984. The interface between semantic roles and initial grammatical relations. In Perlmutter, D. & Rosen, C. (eds.), *Studies in relational grammar*, vol. 2, 38–77. Chicago: Chicago University Press.
Stilo, Donald. 2005. Iranian as buffer zone between the universal typologies of Turkic and Semitic. In Csató, É. Á. & Isaksson, B. & Jahani, C. (eds.), *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic, and Turkic*, 35–63. London/New York: Routledge Curzon.
Stilo, Donald. 2009. Circumpositions as an areal response: The case study of the Iranian zone. *Turkic languages* 13(1). 3–33.
Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of linguistics* 21(2). 385–396.

6
Debonding of inflectional morphology in Kurdish and beyond

Geoffrey Haig

Abstract: The history of case marking across Iranian languages is often described in terms of a grammaticalization cycle, involving the erosion and loss of inherited case markers and their subsequent replacement by innovated case markers via grammaticalization. In this paper I point to certain phenomena in inflectional morphology of Northwest Iranian languages which are difficult to account for within a cyclic view of erosion and replacement. I note unexpected morpheme orderings in Southern and Central Kurdish, and in Gorani (definiteness preceding both case and plural), and the agglutinative nature of the Genitive case in Balochi and Gilaki, both of which are difficult to account for within traditional grammaticalization theory. I conclude that inherited inflectional morphology is not automatically doomed to erosion and loss, but may in fact extend its distributional possibilities and loosen its morphological integration with the base, a process referred to as debonding. I also discuss a cross-linguistically unusual source for the grammaticalization of definiteness marking, which contributes to the unexpected sequences of inflectional morphology in Southern and Central Kurdish, and in Gorani.

1 Identifying the problem

In Northern Kurdish, there is a single overtly-marked case, which is the so-called Oblique. Within Kurdish in its broadest sense (Haig & Öpengin 2014), the Oblique case is also found in Gorani, Zazaki, and in some dialects of Central Kurdish. In Northern Kurdish, there is also a suffix marking indefinite singular, *-ek*, to which the Oblique case marker may attach, as in (1).

(1) N. Kurdish, Muş Dialect

*em ê ji te ra fal-**ek-î** ve-k-in*
1PL FUT ADP 2S.OBL ADP fortune-**INDEF-OBL** PRV-do.PRS.SUBJ-PL

'We will read a fortune for you.' (Haig et al. 2019: muserz01, 0014)

In some varieties of Central Kurdish, and in Gorani, an Oblique case marker may follow a **definiteness** suffix. The latter has approximately the form *-aka*, transcribed *-eke* in some sources (in the examples below, I maintain the transcriptions of the sources cited; the resultant orthographic inconsistencies are irrelevant for the arguments at hand); the initial vowel of the definiteness suffix is generally elided following vowel-final stems. When the noun concerned is plural, the definiteness suffix precedes the plural suffix *-ān*, and the final vowel of the definiteness suffix is generally assimilated to the plural suffix. Examples of definiteness markers followed by a singular Oblique case marker are provided in (2) and (3), while (4) illustrates a definite plural noun:

(2) Central Kurdish, Mukri dialect

*nāme=ī dā be kuř-**eke-ī***
letter=3SG:A give.PST.3SG to boy-**DEF-OBL**

'He gave the letter to the boy.' (Öpengin 2016: 60, glosses adapted)

(3) Hawrami, Luhon dialect

*kɪteb-**aka-y** bāra*
book-**DEF-OBL** bring.IMP.SG

'Bring the book!' (MacKenzie 1966: 16)

(4) Gorani of Gawrajū

*masan pīyā-**k-ān** [...] ẹ̈jāza ni-ma-tī-ya*
for.example man-**DEF-PL** permission NEG-INDIC-give.PRS-3SG

'For example the men [...] don't allow (it).'
(Mahmoudveysi et al. 2012: 143, 7:11)

The morpheme sequence of (in-)definiteness suffix followed by case suffix (1–3), or a plural suffix (4), poses something of a puzzle for historical morphology. If we assume that the indefiniteness marker and the definiteness marker are innovations in these languages (grammaticalized suffixes for indefiniteness and definiteness are not attested in Old Iranian), while the Oblique case suffix and the plural suffix are reflexes of Old Iranian inflectional morphology, then we would expect the older suffixes (case and plural) to occur closer to the base than the supposedly more recent additions (definite and indefinite markers). In other words, we might expect definiteness marking to follow case, or number, as in Swedish, illustrated in (5).

(5)

en	*stol*	*stol-**en***	*stol-ar*	*stol-**ar-na***
INDF	chair	chair-**DEF**	chair-PL	chair-**PL-DEF**
‘a chair’		‘the chair’	‘chairs’	‘the chairs’

But in Kurdish, the order is the other way round, with case and number **outside** of (in-)definiteness. According to a widely-accepted consensus in the literature, the most frequently attested source of definiteness marking is via the grammaticalization of an erstwhile independent deictic element (pronoun or demonstrative; cf. De Mulder & Carlier 2012, and discussion below). The relative position of definiteness marking in Kurdish, however, is hardly compatible with this scenario, because it would imply that the more recently grammaticalized suffixes have somehow intervened between inherited morphology (case and plural) and the base, thus violating the morphological integrity of the word. Furthermore, the order in Kurdish (base-definiteness-number) runs counter to the predictions of Bybee’s Relevance Principle (1985: 13), which suggests that morphology with higher relevance will be closer to the stem. Number has higher relevance than definiteness because it impacts on the notional semantics of the base, while definiteness marking signals information status of an NP in a particular discourse context. The more natural order, at least on the predictions of Bybee (1985), would therefore be base-number-definiteness, as in Swedish (5).

In what follows, I attempt to resolve this puzzle, looking at similar phenomena in other West Iranian languages, and exploring the implications for grammaticalization theory. I will actually suggest two solutions. One is what I term here ‘debonding’, by which I mean that inherited inflectional morphology, in some Iranian languages at least, appears to have weakened its bond with the base and acquired an unusual degree of paradigmatic freedom. Examples of debonding can be found for the category of case (at least

in the singular), and for gender marking. Debonding in itself runs counter to the unidirectionality assumed for much of grammaticalization theory, and is thus of considerable interest. For the Kurdish examples illustrated in (1–4), however, another explanation is more plausible, which does not necessarily involve debonding. Instead I will suggest that the attested morpheme order has arisen through a very unusual grammaticalization path of definiteness morphology in these languages,[1] rather than the debonding of case morphology. It should be emphasized that given the limited time-depth of historical attestation of Kurdish, much of this paper relies on indirect evidence and the conclusions are correspondingly tentative. To my knowledge, the topic has not previously been investigated, and it is hoped that these preliminary thoughts will stimulate further interest. While the main focus of this paper is case morphology, I will also consider examples involving both number, and gender.

The paper is organized as follows. In Section 2, I outline general assumptions on the grammaticalization of case in Iranian, and introduce the notion of debonding. In Section 3, putative examples of debonding are introduced from a range of West Iranian languages. Section 4 takes up the initial puzzle posed by morpheme ordering around the definiteness markers, while Section 5 considers more general conclusions in the light of the data presented.

2 The grammaticalization cycle of case in Iranian

According to Windfuhr (1992), the history of the case system in Iranian can be viewed in terms of a cycle of loss and renewal: the Old Iranian fusional case morphology undergoes phonetic erosion, and ultimately complete loss. The resultant lack of overt case morphology is compensated for by the use of adpositions, which themselves subsequently undergo various processes of phonological reduction and fusion with their complements. Windfuhr (1992: 26) suggests that "each of the many Iranian languages of today represents various stages in that cycle". This view dovetails with widespread assumptions on the grammaticalization of case, which assumes that case affixes develop from erstwhile syntactically independent elements (verbs, nouns, ad-

[1] Since first presenting this paper in 2016, I have become aware of similar processes in other Iranian languages, in particular Balochi (see e.g. Nourzaei et al. 2015). The process, or related ones, have evidently occurred in different Iranian languages beyond Kurdish; this is a topic of ongoing research.

verbial particles, or adpositions), which then fuse with a nominal base to yield case affixes, and ultimately zero (e.g. Heine 2009: 460; Reinöhl 2016).

Although the 'cyclic grammaticalization' approach to the history of case systems remains very influential, a number of phenomena are difficult to reconcile with it. For example, innovated case markers can enter the system **prior** to the complete loss of an inherited system, leading to layered systems. This is widespread in e.g. Indo-Aryan (Reinöhl 2016), or in Nakh-Daghestanian, where various local case suffixes attach to a so-called Oblique stem, often (near) identical with an inherited case marker (e.g. an ergative case in Tabasaran (Comrie & Polinsky 1998), or a genitive in Lezgian (Daniel & Ganenkov 2009: 671)). In other words, the addition of new case markers is not necessarily filling a gap left by the loss of older case marking. Another possibility is that ancient case markers need not erode away to zero, but may in fact "re-vitalize", extending to different host-classes (Haig 2008: 144–152, see also Stilo 2009: 711, and end of Section 3.1 below). None of these developments are readily compatible with the cyclic take on the grammaticalization of case. However, neither do they lend themselves to a uniform explanation, and most researchers are content to provide a taxonomic list of miscellaneous anomalies (e.g. Kulikov 2009). But exceptions to the assumed continuous cycle of loss and renewal are well-attested, and any theory of morphological change surely needs to consider the alternative scenarios (see Kim 2012 for steps in this direction, based on Indo-European).

In this paper I will focus mainly on the fate of the Oblique singular case marker in West Iranian. Although in several languages, including Persian, the history of this case marker accords well with the cyclic view outlined above, in other West Iranian languages, the Oblique case has undergone unpredicted developments. Some of these processes can be seen as examples of **debonding**, following Norde (2009: 186). Debonding involves a weakening of the morphological bond between an inherited inflectional morpheme and its base, and consequently, a gain in prosodic and paradigmatic autonomy of that morpheme. Such a development runs counter to the predictions of grammaticalization theory, which assume that change in inflectional morphemes (there is of course no inevitability in change) will primarily be phonological erosion, and ultimately complete loss. Debonding, on the other hand, yields a case marker that has **greater** positional freedom than its ancestor, and which may spread onto other host categories (e.g. from nouns to pronouns). This kind of development represents a sub-type of those discussed in Norde (2009) under the rubric of degrammaticalization (for more recent discussion of degrammaticalization, see also Ylikoski 2016).

2.1 Inherited versus innovated case

I will assume that in modern Iranian languages, the morphological exponents of structural case can generally be divided into two types: (i) inherited case markers, the reflex of the Old Iranian case inflections; (ii) innovated case markers, historically derived from the grammaticalization of erstwhile adpositions or other items. The distinction was originally sketched in Haig (2008: Ch. 4), but is refined here for the present purposes.

With regard to (i), the inherited case system, the Old Iranian system of inflectional case marking involved complex rules of allomorphy, determined by declensional class, gender, and number. In the transition to Western Middle Iranian, some of the non-nominative cases syncretized, yielding a single marked Oblique case, etymologically a continuation of the old Genitive, but covering a wide range of functions, including adpositional complements, possessors, direct objects, and subjects in the ergative construction (Haig 2008: Ch. 4). In some languages, this Oblique case has been retained, while in others it has disappeared completely, for example in Southern Kurdish, or Persian, which have lost all trace of inherited case marking.[2] In some languages where the inherited Oblique case has been retained, we find a binary opposition between the Oblique, and an unmarked form referred to as the Direct, echoing two-term systems in e.g. Romance (Barðdal & Kulikov 2009), and Indo-Aryan (Reinöhl 2016).

In contemporary Iranian languages, this inherited Oblique case has **approximately** the forms shown in (6). Notably, it generally has distinct forms for singular and plural, and in some languages also for gender (in the singular only), e.g. Northern Kurdish (Haig & Öpengin 2018: 172), or Vafsi (Stilo 2004: 223).

[2] If one analyses the clitic pronouns as suppletive case forms of the free pronouns, then Persian has retained inherited case. In what follows, I will be dealing solely with case marking as it applies to prosodically independent nouns and pronouns, and hence ignoring the clitic pronouns; see e.g. Jügel & Samvelian (2016) for recent discussion of clitic pronouns in contemporary Iranian.

(6) The inherited West Iranian Oblique in the modern languages (nouns; pronouns may differ)[3]

SINGULAR: suffix, consisting of a front, unrounded vowel, phonetically e.g. [-æ, -e, -i, e:, i:]
PLURAL: suffix, consisting of *-ā(n)*, with a rounded vowel in some languages.

A couple of clarifications are in order before we proceed: first, what is referred to here as the 'Oblique' is a morpheme in a given Iranian language for which we assume a common origin (the outcome of syncretisms across the Old Iranian case system). Second, this morpheme may now be associated with different functions in the relevant languages; in other words, 'Oblique' as used here is not a label defined in terms of function, but etymology. Third, descriptions of individual languages may in fact use the label 'Oblique' to refer to a case marker that does not match my usage of the term, rendering comparison across different languages very difficult; see especially the discussion of Tatic and Caspian in Section 3.2.

We turn now to innovated cases. These have a variety of forms, depending on their origins, and may be either en- or proclitics (see Stilo 2009). They can generally be traced to the grammaticalization of syntactically independent items, typically adpositions, or body-part terms like 'head', cf. Gilaki *-sər* 'on(to)' (Rastorgueva et al. 2012: 176). Innovated case markers are an extremely heterogeneous group, so the following definition is couched in negative terms: any case marking used to flag core arguments (direct objects, possessors, or ergative subjects) that cannot be identified with the inherited Oblique, is termed here 'innovated case'.

Over time, innovated case markers may erode to the extent that they become superficially difficult to distinguish from the inherited Oblique. How-

[3] Ignoring the kinship Oblique forms of Middle Iranian (Skjærvø 1983) and their reflexes in e.g. Zazaki (Paul 1998: 22). Likewise, I ignore the suppletive pronominal paradigms here (see Haig 2008: 162–171, 195–197). Note that I have assigned this case system to the earliest form of what is traditionally termed "Western Iranian", although the drawbacks of the traditional East vs. West distinction are well known. Korn (2016) proposes a new sub-grouping of Iranian, involving a previously unidentified group "Central Iranian". The case system that she identifies for "Proto-Central Iranian" (Korn 2016: 421) is basically identical to what I am suggesting here. If Korn's revised classification is validated, this would not affect the argumentation of this paper, except that we could assign languages with vestiges of this assumed inherited case system to the postulated Central Iranian group. Regardless of the outcome of the higher-level classification, there is little doubt regarding the overall shape of the inherited case system.

ever, the phonological properties of innovated markers are not the only clues to their histories. There are in fact distributional diagnostics for distinguishing inherited from innovated case. These are summed up in Table 1 below. For ease of exposition, I have taken the inherited Oblique of Northern Kurdish Oblique (*-ī /-ē/-ān*) as a typical example of inherited Oblique, while the Persian accusative clitic *=rā* illustrates a typical innovated case marker.

Having spelled out the fundamental difference between inherited and innovated case, we can now proceed to some examples of debonding of inherited case, and related phenomena in a selection of West Iranian languages.

Table 1: Diagnostics for distinguishing innovated case markers from the inherited Oblique in West Iranian

Criterion	Inherited Oblique	Innovated Case
Allomorphy	Allomorphy depending on class of base: plural bases have different Obl. to singular; masculine and feminine singular Obl. may differ; pronouns may have suppletive rather than suffixal Obl.; in some dialects of Northern Kurdish masculine singular Obl. is expressed through Ablaut rather than suffix (Haig & Öpengin 2018: 209–210)	Uniform exponence (barring low-level phonetic processes), i.e. single form across different classes of base
Bondedness	Affix (inseparable from noun, but see Section 3 below), may assimilate stem-final vowels (e.g. Northern Kurdish *li ser masê* < *mase+ê* 'on the table-OBL')	Clitic (phrasal rather than nominal host, freedom of host selection)
Fusion with other inflectional categories	Yes (with plural number, see allomorphy above)	No
Sensitive to information structure	Uncommon	Very common with innovated case on direct objects, which generally exhibits DOM
May be added to an existing inherited Oblique	No	Yes (in other Iranian languages at least, though not in Persian)
Suspended affixation in coordinated NPs	Uncommon, but see below	Yes

3 Debonding in West Iranian

3.1 The Genitive case marker in Balochi and Gilaki

Balochi is a geographically and structurally very diverse group within northwestern Iranian. The case system is more complex than in Kurdish, and different researchers have adopted different analyses. As Jahani & Korn (2009: 651) point out, there is "no agreement in grammatical descriptions of Balochi on the number of cases and what they should be called [...]". In this section I will briefly consider the nature of the Genitive case in Balochi, a suffix which can be identified across most dialects of Balochi, and is variously transcribed with *-ī*, *-ē*, *-e*, *-a*, or *-ay* depending on the dialect, and the source (Jahani & Korn 2009: 651; Nourzaei 2017: 37–38, 43, 55, 61). Given the contentious nature of the case system, and the range of dialectal variation across Balochi, my comments remain tentative. However, I believe the Genitive in Balochi provides us with a potentially fruitful window on the nature of historical change in the case systems of Iranian. I will also discuss a very similar suffix in another northwest Iranian language, Gilaki, and in Tatic, which I believe shed further light on the history of this case.

Throughout Balochi, the main function of the Genitive is to mark prenominal possessors, as in (7).

(7) Balochi of Turkmenistan

*gis-**ay*** *wāund mēmān-**ay*** *abar-ā* *uškit*
house-**GEN** owner guest-**GEN** word-OBJ hear.PST.3SG

'The owner **of the house** heard the words **of the guest**.'
(Axenov 2006: 79)

The origin of this suffix remains to be established with certainty. One possibility is that it is a continuation of the inherited Oblique discussed above. In favour of this position we can note that the modern Balochi Genitive is usually the case assigned by postpositions, e.g. *ē mulk-**ay** tā* 'this country-GEN in = in this country' (Sistani Balochi, Delforooz 2010: 151). Furthermore, it can occur between a noun or pronoun and another case marker, variously referred to as Oblique II, or Locative (Jahani & Korn 2009: 652). It also exhibits some allomorphy, for example in Sistani and Koroshi Balochi, it has the form *-ī* after pronouns, but *-ay* after nouns (Nourzaei 2017: 55); a similar pattern obtains for Balochi of Turkmenistan, where the genitive marker is *-ī* with pronouns

and proper nouns, and after the plural suffix, but *-ay* for other nouns (Axenov 2006: 72). These facts lend support to an analysis as an inherited case marker, and a connection to the inherited singular Oblique discussed in the preceding section. We can also note that another Iranian language, Bashkardi, has a sporadically-used *-ī* suffix, which also marks possessors. Korn (2017: 11) considers the Old Iranian genitive (the primary source of the inherited Oblique) to be a possible candidate for the origins of the Bashkardi suffix; see also the discussion on Tatic below. Finally, I am unaware of any more convincing alternative explanation (i.e. some kind of innovated case marker, from an earlier postposition?) that would account for both the form of the Balochi Genitive, some of its distributional properties, and the evident commonalities across the otherwise fairly diverse group of Balochi dialects.[4] My provisional conclusion, then, is that the Balochi Genitive is cognate with the singular form of the inherited Oblique in other West Iranian languages.

However, the Balochi Genitive also exhibits a number of properties that are difficult to reconcile with the inherited singular Oblique case. First of all, as already mentioned above, it occurs outside of the plural marker, i.e. in a position where the singular Oblique was historically never attested, see (8).

(8) Balochi of Sistan

ges-ān-ī *dapā*
house-**PL-GEN** in_front_of

'in front **of the houses**' (Nourzaei 2017: 690)

Not only does it occur outside the inherited plural marker *-ān*, widely attested throughout west Iranian, but it also occurs outside of the innovated plural suffix *-obār*, which is found in Koroshi Balochi. I am unaware of any convincing etymology for this suffix, which appears to be unique to Koroshi, and is presumably an innovation:

[4]One difficulty remains with the claim of a historical link to the inherited Oblique, namely the fact that the Balochi Genitive is not used to mark the subject of an ergative construction, whereas in other Iranian languages that have maintained the inherited Oblique, it is typically the case used for this function. However, examples of innovated case, rather than Oblique case, being used for the subjects of ergative constructions are attested elsewhere in Iranian, see Haig (2008: 167) for discussion.

(9) Koroshi Balochi

negahbān-obār-ay *basāb jāh xālī bod-a*
watchman-PL-GEN actually place empty become.PST.3SG

'You know, **the watchmen's** place was empty [...]'
(Nourzaei 2017: 630)

Again, this is difficult to reconcile with the view that the Genitive is an inherited case marker. Furthermore, the Genitive case is also regularly attached to personal pronouns, and to the reflexive pronoun, e.g. Coastal Balochi *man-ī* '1SG-GEN', *ta-ī* '2SG-GEN', etc. (Nourzaei 2017: 55), or Sistani Balochi *wat-ī* 'REFL-GEN' (Delforooz 2010: 221).

Another northwest Iranian language, Gilaki, also has a case labeled "Genitive" (Rastorgueva et al. 2012: 56–57), with the form *-ə*.[5] As in Balochi, the Gilaki Genitive marks adnominal possessors, and the complements of postpositions. Both functions are illustrated in (10), where the Genitive is glossed as a clitic, following the conventions of the source:

(10) Gilaki

*ašk məryəm=**ə** čašm=**ə** dor [...]*
tear Maryam=**GEN** eye=**GEN** around [...]

'Tears [...] around Maryam's eyes'
(Rastorgueva et al. 2012: 420, glosses as in original)

The Gilaki Genitive is regularly attached outside the plural suffix of a noun (*gul-ón-ə* 'flower-PL-GEN', Stilo 2018: 692, Table 5A; Rastorgueva et al. 2012: 56) and also to a plural pronoun, as in the following. Note that the form of the demonstrative is paralleled in Balochi, where the demonstrative pronoun can likewise take the Genitive case, cf. Balochi *ēšān-ī* DEM.PL-GEN (Korn 2005: 334).

[5] The Gilaki genitive takes the form *-i* with pronominal stems (Rastorgueva et al. 2012: 91), and *-e* in some other environments, underscoring the similarity to the Genitive in Balochi. These details of allomorphy can hardly be coincidence.

(11) Gilaki

ušán-ə *xånə šimi [...]*
DEM.PL-GEN house go.PRS.1PL

'(that) we are going to **their** house [...]'
(Rastorgueva et al. 2012: 226, glosses supplied)

Rastorgueva et al. (2012: 56, fn. 3) claim that the genitive is etymologically the continuation of the Old Iranian Genitive, i.e. an inherited Oblique case. As in Balochi, in Gilaki too the Genitive case marker also precedes postpositions, including benefactive *=re* (Rastorgueva et al. 2012: 64).

It seems reasonable to assume that the Genitive of Gilaki and Balochi are cognate,[6] but whether they should be considered reflexes of an inherited Oblique, or of some as yet unknown shared innovation in Gilaki and Balochi, is not yet settled. Suggestive evidence in favour of the former position is available from another group of northwest Iranian languages, Tatic, which I will briefly sketch here. In Tatic, just like Balochi and Gilaki, we find prenominal possessors marked by a suffix variously transcribed with *-i*, *-í*, and *-e* in Stilo (2018: 698–699), e.g. *hæsæn-i kætæb* 'Hasan's book' (Harzani dialect).[7] In one sub-group of Tatic languages (including e.g. Vafsi), the form of this case suffix varies in the singular according to the gender of the noun (Stilo 2018: 694), while in other Tatic languages (including the Talyshi group) gender is lost, and this suffix has an invariable form in the singular. In some dialects of Tatic, there is a suffix that marks pre-nominal possessors, which may also occur after the plural suffix, and with pronouns. This is illustrated in (12a–12c):

[6]Note that the assumed shared proto-morpheme would have had at least two distinct allomorphs, depending on the gender of the noun to which it attached. It is therefore possible that the daughter languages may have continued either of the two allomorphs, making it difficult to regularly reconstruct a single phonological form for the suffix (in this case: a particular vowel quality). My suggestion of cognacy is therefore primarily based on a very gross measure of phonological similarity, and the overwhelming functional and distributional equivalence.

[7]Comparison across Gilaki (part of Stilo's (2018) Caspian group), and Tatic, are rendered more difficult due to differences in terminology. What Rastorgueva et al. (2012) refer to as the "Genitive" in Gilaki is called by Stilo the "Possessive". What I consider etymologically the same morpheme in Tatic is referred to as the "singular Oblique", or the "Reverse Ezafe" in Stilo (2018).

(12) Tāleqāni: Orāzāni dialect of Tatic

a. *pa-i* *bon*
foot-GEN bottom
'sole of the foot'

b. *boz-an-i* *šir*
goat-PL-GEN milk
'the goats' milk'

c. *mən-i* *xanæ*
1SG-GEN house
'my house' (stressed vowel in *mən*)

(Stilo 2018: 700, glosses modified)

Stilo (2018: 700) glosses this suffix as the "Reverse Ezafe", which he considers distinct from the Oblique suffix of Tatic. My own view is that it is the reflex of the old Oblique, which has lost gender differentiation, yielding a uniform phonetic form. Consequently it may debond, yielding exactly the same kinds of morpheme sequences that are attested in Gilaki and Balochi. As part of this general re-structuring of nominal inflection, the plural suffix becomes a general marker of plural number (rather than a composite morpheme expressing Oblique and plural), and then permits combinations with the debonded Oblique marker.

To return now to our point of departure, the Genitive suffix in Balochi and Gilaki, we note that it can be indiscriminately attached to hosts of different categories, and follow other inflectional morphology in a manner typical for agglutinative morphology. This is not what we expect of an inherited case marker in an Indo-European language. I have nevertheless argued that the source of the Genitive in these languages is an inherited case marker, rather than an innovation. If these claims are correct, then we need to accept that inherited case morphology (or more generally, inflectional morphology) is not necessarily doomed to erode to zero. Rather, under conditions which are yet to be specified, an inherited case suffix may actually increase in productivity, extending to additional form classes with which it was previously not associated (e.g. nouns to pronouns, or from singular nouns to plural nouns). When this happens, we find morpheme sequences of e.g. PLURAL-CASE, which are historically unattested with the original morpheme.[8] This is

[8] It is worth noting that it is also the genitive case in Germanic which has undergone unexpected developments with certain parallels to those outlined here for Balochi and Tatic (see Norde

presumably linked to some minimal threshold of phonological saliency and uniformity of the original morpheme, and perhaps to the nature of the functions with which the original morpheme is associated; this requires more research. I refer to this process as debonding, following Norde (2009), but other terms would be equally appropriate. Stilo (2009: 711), for example, refers to a process of "agglutinative analogy" in connection with the Balochi Genitive, which aptly highlights the nature of the resulting structures. Once gender distinctions are lost, leaving a uniform oblique singular suffix, it seems that it may be interpreted as an all-purpose genitive marker, and attached to, for example, plural marked nouns or pronouns. Interestingly, this process also seems to depend on the presence of some form of innovated case marker that takes over the direct object function of the inherited Oblique; this requires more research.

3.2 Related examples of debonding

Another example of debonding of case morphology comes from the Tatic dialect of Dikin Marāqei of Alamut. Here, the Oblique suffix may follow a pronominal clitic:

(13) Tatic of Dikin Marāqei of Alamut

*sær=**t-i***	*me-jæn-én*
head=**2SG.POSS-OBL.M**	ASP-hit-1SG

'I'll hit your head (m.)' (Stilo 2016 and personal communication)

However, it is only the **masculine singular Oblique**, and possibly the feminine Direct suffix (of uncertain origin), which allow displacement as in (13). With the feminine singular Oblique (going back to an old kinship Oblique), and with the plural Oblique, the possessive clitic occurs outside the case marker. This is shown in (14), where the feminine Oblique marker precedes the possessive clitic:

2009, among others). Whether this is pure coincidence, or whether genitive cases are generally more prone to debonding than other case markers remains an open question.

(14) Tati, Dikin Marāqei of Alamut

*æz das-**ær=et** mi-n-ɪn*
1SG hand-**OBL.F=2SG.POSS** TAM-see.PRS-1SG

'I see your hand (f.)' (Stilo 2016 and personal communication)

A related phenomenon occurs in Gorani when the Oblique suffix attaches to a phrasal (NP) host, rather than a lexical one (N). Thus the marker concerned is no longer a nominal, but a phrasal affix, showing the typical distributional properties of an innovated rather than an inherited case marker. In (15), it occurs on an adjectival host, presumably because this is the rightmost boundary of the NP (and in fact the same phenomenon occurs with the definiteness suffix in this example):

(15) Awroman dialect of Gorani

a. *kitéb-aká*
book-DEF
'the book'

b. *kitéb-a siāw-aká*
book-IZ black-DEF
'the black book'

c. *[kitéb-a siāw-aká]-y*
book-IZ black-DEF-**OBL**
'the black book' (direct object)

(MacKenzie 1966: 17–18, cited in Haig 2008: 145–146, transcription follows original)

The final example comes from gender morphology in Tatic. In a number of Iranian languages, gender is maintained as a grammatical category, systematically reflected in various parts of the grammar. As far as I can ascertain, gender in western Iranian is always a two-term system (traditionally labeled masculine and feminine), and gender is only relevant in the singular; gender distinctions are neutralized in the plural. It seems evident that gender in Iranian, where it is found, is an inheritance from Old Iranian, rather than an innovation.

In Tatic languages, gender is found in about half of the group (Stilo 2018). A fairly typical system is shown in Table 2.

Table 2: Gender marking on nouns, Karani dialect of Tatic (Yarshater 2009: 555)

		Masculine	Feminine
Singular	DIRECT	*-ø*	*-a* (unstressed)
	OBLIQUE		*-e* (stressed)
Plural	DIRECT		*-e* (stressed)
	OBLIQUE		*-ān* (stressed)

Gender is also reflected in agreement with certain kinds of predicate, for example the present tense of the copula:

(16) Karani dialect of Tatic

a. *Hasan dalú-e*
Hasan crazy-COP.PRS.MASC
'Hasan is crazy (m.).'

b. *Zeynab-a dalu-ā*
Zeynab-FEM crazy-COP.PRS.FEM
'Zeynab is crazy.' (Yarshater 2009: 555, glosses added)

Remarkably, the feminine singular direct suffix *-a* on a noun can be separated from its base by a clitic, i.e. can be debonded. This happens in clauses containing a past transitive verb, when a clitic pronoun indexes the transitive subject, and the rules of clitic placement conspire to leave a feminine singular noun as the landing site for the clitic. This is shown in (17), where the third singular subject-indexing clitic is *=eš*, which separates the feminine suffix from the direct object 'wild-goat'. The feminine gender of the direct object is indexed on both verbs by a corresponding agreement suffix; in the second clause, the feminine agreement marker on the verb is also displaced by a clitic pronoun:

(17) Karani dialect of Tatic

*em naccira_bez=**eš-a** bezzi-ā bard=eš-ā de:*
this wild_goat=**3SG-FEM** shoot.PST-FEM bring.PST=**3SG-FEM** village

'He shot (f.) this **wild goat** (f.) and brought (f.) to the village.'
(Yarshater 2009: 555, glosses added; see also Yarshater 2009: 565 for further examples)

See also Öpengin (this volume) for examples from Kurdish of clitic pronouns indexing a past transitive subject, and displacing what appears to be an inflectional suffix. It is worth noting that the so-called 'feminine suffix' that we have considered here has undergone a functional shift from a gender marker towards becoming a marker of definiteness and individuation. To what extent this functional change can be linked to its debonding, is an open question.

4 The case-after-definiteness puzzle

Let us return now to the question posed at the outset of this paper, namely how indefiniteness and definiteness markers in Kurdish, presumably innovations, should nonetheless occur inside inherited case and number morphology, cf. examples (1–4) above. Given what I have claimed above regarding debonding of inherited morphology, an obvious solution would be to assume a debonding scenario in Kurdish, leading to a loosening of the bond between case-suffix and base and the possibility of morpheme re-ordering. However, there are reasons why this may not be the correct solution, at least with regard to the relative positioning of the definiteness suffix.

If we consider first the indefinite suffix, we know that a suffix for singular indefinites is extremely widespread throughout West Iranian, often alternating, or combining, with a pre-posed indefinite article, generally transparently related to the numeral 'one'. The Kurmanji indefinite suffix *-ek* appears superficially to be related to the numeral *yek* 'one', a typical source of indefinite singular markers cross-linguistically. If that is the source for the Kurmanji indefiniteness suffix, then it is presumably a later development, which occurred after the Old Iranian period. On that assumption, the ordering of the Oblique suffix outside the indefiniteness suffix would be a clear example of case debonding. However, given the widespread presence of indefiniteness suffixes in West Iranian (in various forms), it is not impossible that the Kurmanji indefiniteness marker could be an older layer of morphology, rather than the product of later grammaticalization. A second possibility would be that the suffix *-ek* arose through the univerbation of, perhaps, an appositional postposed *yek*, which could have been Oblique-marked itself, and later fused with the noun. But as long as the history of the indefiniteness suffix itself remains unresolved, it is difficult to decide on the source of this morpheme sequence.

With regard to the definiteness suffixes of central and southern Kurdish, and Gorani, the problem is somewhat different. Systematic marking of definiteness is not historically attested in Old Iranian. Within Kurdish, it is only found in Central and Southern Kurdish, but not in Northern Kurdish. An initial assumption would be, then, that it represents an innovation in Central and Southern Kurdish. Cross-linguistically, the grammaticalization of definiteness markers is one of the best-known topics in grammaticalization theory. The widely-cited case-studies involve an origin from some kind of independent deictic element (often a demonstrative or pronominal element) that loses deictic force and prosodic independence, finally becoming a general marker of discourse identifiability (De Mulder & Carlier 2012; Himmelmann 2001). If the same kind of development were behind the Central Kurdish definiteness marker *-aka*, then the position of Oblique case and plural number **outside** of this suffix would be difficult to account for without assuming debonding of case and number morphology.

For Central Kurdish *-aka*, however, there is no obvious demonstrative or pronominal element that could have provided the historical source for such a grammaticalization. Furthermore, it turns out that there are candidate suffixes attested in Old and Middle Iranian that might have provided the source.[9] If these are indeed the origins of the definiteness marker, then we are not dealing with a typical example of the grammaticalization of definiteness (e.g. demonstrative to definiteness marker), but rather with some kind of re-analysis or functional shift of existing morphology to yield a definiteness suffix. A detailed investigation of the origins of the definiteness suffix in Kurdish is beyond the scope of this paper (and is the topic of ongoing research). The following remarks nevertheless provide a working hypothesis for explaining the otherwise puzzling ordering of definiteness inside of case and number in Central and Southern Kurdish, and in Gorani.

For Old Iranian, Ciancaglini (2012) discusses the reflexes of an Indo-European derivational suffix **-ko-*, which yielded **-ka-* in Old Iranian. It seems to have been remarkably productive and could apply to a variety of bases, including personal pronouns. Its semantics were vague, but included relationality (e.g. with pronouns: 'you' > 'yours'), but also a diminutive sense, or one of endearment. Ciancaglini (2012: 92) also notes that it occurs in contexts

[9] I am extremely grateful to Johnny Cheung for bringing these possibilities to my attention, and to Thomas Jügel for providing additional references and material. Neither bear any responsibility for my interpretation of this material. A reviewer drew my attention to Jahani (2015), who independently suggests the same development for Central Kurdish, and indeed for the colloquial Persian definiteness marker *-e*.

where it appears to add no particular semantic content at all. In Old Persian, the main attested functions are in combination with thematized stems, and the resulting meaning is something like 'the one characterized by X', thus Old Persian **banda* 'bond, fetter' > *bandaka* 'subject, vassal, servant'. She further notes that it is very frequent with "toponyms and ethnonyms designating non-Iranian peoples, or peoples geographically distant or little known to the Persians" (2012: 95). Interestingly, in this context the suffix appears to add nothing to the denotational semantics, but fulfills evidently some kind of emphatic or contextually determined role. In Avestan, and in other ancient Indo-European languages, words with this suffix are often linked to informal registers, occurring in "imprecatory, pejorative, or affective and familiar contexts" (Ciancaglini 2012: 95). This may explain why this kind of usage is concentrated in Young Avestan, but is scarce in the Gāthās and the Old Persian inscriptions, with their more formal and ritualized character.[10]

For western Middle Iranian, a suffix *-ag* is noted as "one of the most productive" suffixes (Durkin-Meisterernst 2014: 155). It is evidently related to the *-ka/-aka* complex of Old Iranian just discussed, and has a similarly varied functional spectrum. It could create adjectives, or add a sense of diminution or endearment to a noun, among diverse other functions noted by Durkin-Meisterernst (2014: 156–158). It also occurred with phonological variants *-ak* or *-k*.

Although we lack historical records of the direct Middle Iranian predecessors of Central Kurdish, it seems reasonable to assume that an inherited suffix, with this degree of productivity in Parthian and Middle Persian, would also have been present in the precursors of Kurdish. The more pressing question is whether a suffix of this nature could have developed into a marker of definiteness? Within the grammaticalization literature, diminutives have not figured as possible sources of definiteness markers, but recently Pakendorf & Krivoshapkina (2014) point to an interesting parallel in Ėven, a Tungusic language of Siberia. The authors note that suffixes with evaluative semantics (traditionally termed 'diminutives') have developed into markers of discourse identifiability, approximately comparable with the function of definite articles in languages of northwestern Europe. A recent cross-linguistic survey of diminutives (Ponsonnet 2018) points to the broad range of semantic types regularly associated with so-called diminutives. It is clearly not the

[10] However, some of the examples noted by Ciancaglini are arguably reflexes of a deverbal suffix **-aka*, or result from a thematic *-a* attached to a stem-final *-k*. In other words, we must reckon with a certain amount of opacity and reanalysis in the interpretation of the relevant forms.

case that diminutives are primarily markers of 'small size'; rather, they regularly express speakers' subjective evaluations, often indicating endearment and familiarity, but also ridicule or contempt. From this perspective, a development along the lines of endearment > familiarity > identifiability, definiteness, as suggested in Haig (2018a), does not seem implausible. A link from inherited diminutive marking to definiteness marking has been suggested for northwest Iranian Balochi (Koroshi dialect, Nourzaei et al. 2015: 32).

Whether the Central and Southern Kurdish definiteness suffix[11] can indeed be traced back to the Middle and Old Iranian 'diminutives' remains an open question. But it nevertheless remains a plausible theory that would neatly account for the suffix ordering puzzle outlined at the outset of this section. The present account assumes that definiteness in Kurdish arises not through the grammaticalization of some previously independent morpheme, but via excrescence: a former derivational suffix (with vague and as yet not fully understood semantics, but apparently involving endearment and familiarity) is re-analysed to become an inflectional suffix indicating information status of the noun phrase concerned. Given that the suffix was historically part of the base, then the current position of case markers outside of this suffix is quite natural, and we need not invoke a process of case debonding. The same applies to the combination of definiteness suffix and plural marking (4), which again would reflect the historical sequence of these suffixes, rather than any kind of debonding.

Thus, for the combination of definiteness and case, what appears at first glance to be the result of debonding in may in fact have a different historical explanation, namely the development of the definiteness marker from an old derivational suffix. This possibility has largely remained obscured due to the lack of comparable examples in the grammaticalization literature on definiteness.

[11] It needs to be pointed out that the Central Kurdish "definiteness" suffix is not functionally equivalent with the English definite article, despite the recent claims to this effect in Zahedi & Mehrazmay (2011). The naturalistic texts available for Central Kurdish include numerous referential NP's with unambiguously discourse-identifiable referents, which translationally would require a definite article in English, yet which lack the so-called definiteness suffix. On the assumption that the source morpheme was a diminutive, this kind of "optional definiteness" is actually not surprising, see Haig (2018a).

5 Summary and outlook

The point of departure for this inquiry was the unusual sequence of definiteness marking inside of the Oblique case and number morphology in Central and Southern Kurdish, and in Gorani. This led to an investigation of the history of case marking in west Iranian, and in particular, to the history of the inherited Oblique case marker (most likely the continuation of the Old Iranian Genitive), and to a more general considerations regarding the processes of morphological change. I will first summarize the main conclusions regarding the history of case marking in Iranian, before taking up some of the broader issues at the end of this section.

According to widely-held views on the history of inherited case marking in much of Indo-European (Kim 2012), the expectation is that it gradually erodes, finally yielding zero, and may then be subsequently replaced by various kinds of innovated case markers through the process of grammaticalization. Any comparative discussion of case marking in Iranian therefore needs, at least in principle, to distinguish between the exponents of inherited case on the one hand, and innovated case markers on the other. In Section 2.1, I laid down a set of criteria for this purpose, and for many of the case markers of Iranian, the distinction can readily be maintained. I then went on to investigate the Genitive suffix in Balochi and Gilaki, which looks in many respects like an inherited case marker (and some authors have claimed it is), yet it exhibits agglutinative distributional properties that were not attested in its assumed ancestor morphemes, and which would be more typical of innovated case marking. I investigated the assumed cognates of this case marker in Tatic, where in some dialects it exhibits properties more or less typical of inherited Oblique case (gender and number determined allomorphy, use in other functions outside of adnominal Genitives), while in others it resembles the Genitive of Gilaki and Balochi. My suggestion is that in origin, the Genitive of Balochi and Gilaki is indeed the inherited Oblique, but it has debonded, and extended to additional environments where its ancestor was never attested. If this analysis is correct, then the history of inherited case is not a one-way path to zero. Instead, case markers may be involved in paradigm restructuring, loosen their bond to the base, and extend to related form classes (nouns to pronouns, for example). The result of this process actually ends up looking more like an innovated case marker than an inherited one.

The odd ordering of case and definiteness in Southern and Central Kurdish could in principle also have involved debonding of case marking. However, I

suggest that the morpheme ordering puzzle in these languages is related to the history of the definiteness suffix, rather than to the case suffix. My suggestion is that the definiteness suffix goes back to an old derivational morpheme, traditionally (and somewhat misleadingly) termed a 'diminutive'. On that account, then the attested morpheme order is quite natural. But notice again how this account involves a process of reanalysis and extension of an existing suffix, rather than grammaticalization of previously independent lexical, or less grammatical material.

I also noted further examples of presumably inherited morphology showing symptoms of debonding, for example the separation of case marking (14), or gender marking (17) from the noun stem in Tatic. Notably, both these instances involve the intrusion of a clitic between inflectional morphology and stem; a similar kind of phenomenon is attested with verbal morphology in Central Kurdish, see Öpengin (this volume). While these may appear to be fairly isolated, they do raise serious questions regarding our understanding of morphological change, and the adequacy of the grammaticalization paradigm for addressing them. Similar examples of paradigm restructuring and replacement processes within inflectional morphology have been discussed by several authors, drawing on data from a wide range of languages (e.g. Janda (1996) on Slavic, Heath (1997) on languages of Australia; see also Reinöhl & Himmelmann (2017) for critical discussion). Willis (2016) is an attempt to unite some of these observations into a more coherent theory, drawing on the assumed obsolescence of the morphology concerned, and the process of exaptation. However, I am not fully convinced that this can be applied to the Iranian examples, because it is not at all evident that, for example, the inherited Oblique case marker is devoid of function. For the time being, I will simply note that the familiar clines of grammaticalization do not simply run on seamlessly into the realm of inflectional morphology; the creation and organization of inflectional morphology do not arise through the continued application of the forces of grammaticalization, see Haig (2018b: 813) and also Enger (2013) on the autonomy of morphology.

In closing, two final points are noteworthy. First, in several northwest Iranian languages, including Northern and Central Kurdish, Balochi, Gilaki and generally Caspian, we find the emergence of agglutinative structures in nominal morphology, involving the categories of case, number and definiteness. This is a major change when compared to Old and Middle Iranian languages, where inflectional morphology tended to be fusional, rather than agglutinative. This is presumably one of the outcomes of debonding. Viti (2015) points to a predominance of agglutinative structures among the attested cases of

degrammaticalization, and this would tie in well with the overall findings here. Second, the inherited Oblique case has undergone very divergent developments across West Iranian: (i) complete loss, without replacement, as in Southern Kurdish; (ii) complete loss, but with replacement via innovated case (e.g. Persian); (iii) maintenance (Northern Kurdish, Zazaki); maintenance, with debonding and extension (Balochi, Gilaki?). These variant outcomes may be a reflection of the status of case as an intermediate category in the scale of the retention of morphology identified by Roberts & Bresnan (2008), who locate case between number (most likely to be retained) and gender (most likely to be lost). This would also fit well with what is known regarding these two categories in Kurdish: inherited plural morphology is generally retained, and if it is lost, it is always replaced by innovative morphology (cf. the Balochi plural in (9) above); inherited gender morphology is sometimes retained, and often lost; if lost, it is never replaced.

References

Axenov, Serge. 2006. *The Balochi language of Turkmenistan: A corpus-based grammatical description*. Uppsala: Acta Universitatis Upsaliensis.

Barðdal, Jóhanna & Kulikov, Leonid. 2009. Case in decline. In Malchukov, A. & Spencer, A. (eds.), *The Oxford handbook of case*, 470–478. Oxford: Oxford University Press.

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.

Ciancaglini, Claudia. 2012. Outcomes of the Indo-Iranian suffix **-ka* in Old Persian and Avestan. In Basello, G. P. & Rossi, A. V. (eds.), *Dariosh studies II: Persepolis and its settlements. Territorial system and ideology in the Achaemenid state*, 91–100. Napoli: L'Orientale, Dipartimento Asia, Africa e Mediterraneo.

Comrie, Bernard & Polinsky, Maria. 1998. The great Daghestanian case hoax. In Siewierska, A. & Song, J. J. (eds.), *Case, typology and grammar*, 95–114. Amsterdam: Benjamins.

Daniel, Michael & Ganenkov, Dmitry. 2009. Case marking in Daghestanian: Limits of elaboration. In Malchukov, A. & Spencer, A. (eds.), *The Oxford handbook of case*, 668–685. Oxford: Oxford University Press.

De Mulder, Walter & Carlier, Anne. 2012. The grammaticalization of definite articles. In Heine, B. & Narrog, H. (eds.), *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press. (`https://doi.org/10.`

1093/oxfordhb/9780199586783.013.0042). (Accessed 2017-02-28). [Online publication date Sep. 2012].

Delforooz, Behrooz Barjasteh. 2010. *Discourse features in Balochi of Sistan*. Uppsala: Acta Universitatis Upsaliensis.

Durkin-Meisterernst, Desmond. 2014. *Grammatik des Westmitteliranischen (Parthisch und Mittelpersisch)*. Wien: Verlag der Österreichischen Akademie der Wissenschaft.

Enger, Hans-Olav. 2013. Inflectional change, 'sound laws' and the autonomy of morphology: The case of Scandinavian case and gender reduction. *Diachronica* 30(1). 1–26.

Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A Construction Grammar approach*. Berlin: Mouton de Gruyter.

Haig, Geoffrey. 2018a. Optional definiteness in Central Kurdish and Balochi: Conceptual and empirical issues. Invited talk at the *Third Workshop on Information Structure in Spoken Language Corpora (ISSLaC3)*, University of Münster, 7–8 December 2018.

Haig, Geoffrey. 2018b. The grammaticalization of object pronouns: Why differential object indexing is an attractor state. *Linguistics* 56(4). 781–818.

Haig, Geoffrey & Öpengin, Ergin. 2014. Kurdish: A critical research overview. *Kurdish studies* 2(2) [Special issue: *Kurdish linguistics: Variation and change*]. 99–122.

Haig, Geoffrey & Öpengin, Ergin. 2018. Kurmanji Kurdish in Turkey: Structure, varieties and status. In Bulut, C. (ed.), *Linguistic minorities in Turkey and Turkic-speaking minorities of the peripheries*, 157–229. Wiesbaden: Harrassowitz.

Haig, Geoffrey & Vollmer, Maria & Thiele, Hanna. 2019. Multi-CAST Northern Kurdish. In Haig, G. & Schnell, S. (eds.), *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts*. (https://multicast.aspra.uni-bamberg.de//#nkurd) (Accessed 2019-09-05).

Heath, Jeffrey. 1997. Lost wax: Abrupt replacement of key morphemes in Australian agreement complexes. *Diachronica* 14(2). 197–232.

Heine, Bernd. 2009. The grammaticalization of cases. In Malchukov, A. & Spencer, A. (eds.), *The Oxford handbook of case*, 458–469. Oxford: Oxford University Press.

Himmelmann, Nikolaus. 2001. Articles. In Haspelmath, M. & König, E. & Oesterreicher, W. & Raible, W. (eds.), *Language typology and language universals: An international handbook*, vol. 1, 831–841. Berlin: De Gruyter.

Jahani, Carina. 2015. On the definite marker in Modern Spoken Persian. Paper presented at the *Sixth International Conference on Iranian Linguistics (ICIL6)*,

23–26 June 2015, Ilia State University, G. Tsereteli Institute of Oriental Studies, Tbilisi/Georgia.

Jahani, Carina & Korn, Agnes. 2009. Balochi. In Windfuhr, G. (ed.), *The Iranian languages*, 634–692. London: Routledge.

Janda, Laura. 1996. *Back from the brink: A study of how relic forms in languages serve as source material for analogical extension*. Munich/Newcastle: Lincom Europe.

Jügel, Thomas & Samvelian, Pollet. 2016. Les pronoms enclitiques dans les langues ouest-iraniennes: Fonctions et distribution géographique. *Bulletin de la Société de Linguistique de Paris* CXI(1). 391–432.

Kim, Ronald. 2012. The Indo-European, Anatolian, and Tocharian "secondary"cases in typological perspective. In Cooper, A. I. & Rau, J. & Weiss, M. (eds.), *Multi nominis grammaticus: Studies in classical and Indo-European linguistics in honor of Alan. J. Nussbaum on the occasion of his sixty-fifth birthday*, 121–142. Ann Arbor/New York: Beech Stave.

Korn, Agnes. 2005. *Towards a historical grammar of Balochi: Studies in Balochi historical phonology and vocabulary*. Wiesbaden: Reichert.

Korn, Agnes. 2016. A partial tree of Central Iranian: A new look at Iranian subphyla. *Indogermanische Forschungen* 121(1). 401–434.

Korn, Agnes. 2017. Notes on the nominal system of Bashkardi. *Transactions of the Philological Society* 115(1). 79–97.

Kulikov, Leonid. 2009. Evolution of case systems. In Malchukov, A. & Spencer, A. (eds.), *The Oxford handbook of case*, 439–457. Oxford: Oxford University Press.

MacKenzie, David. 1966. *The dialect of Awroman (Hawrāmān-ī Luhōn): Grammatical sketch, texts, and vocabulary*. København: Munksgaard.

Mahmoudveysi, Parvin & Bailey, Denise & Paul, Ludwig & Haig, Geoffrey. 2012. *The Gorani language of Gawraju (Gawrajuyi), a village of west Iran: Grammar, texts and lexicon*. Wiesbaden: Reichert.

Norde, Rachel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

Nourzaei, Maryam. 2017. *Participant reference in three Balochi dialects: Male and female narrations of folktales and biographical tales*. Uppsala: Acta Universitatis Upsaliensis.

Nourzaei, Maryam & Jahani, Carina & Anonby, Erik & Ahangar, Abbas Ali. 2015. *Koroshi: A corpus-based grammatical description*. Uppsala: Acta Universitatis Upsaliensis.

Öpengin, Ergin. 2016. *The Mukri variety of Central Kurdish: Grammar, texts, and lexicon*. Wiesbaden: Reichert.

Pakendorf, Brigitte & Krivoshapkina, Ija. 2014. Even evaluatives and the marking of definiteness. *Language typology* 18(2). 289–331.

Paul, Ludwig. 1998. *Zazaki: Grammatik und Versuch einer Dialektologie*. Wiesbaden: Reichert.

Ponsonnet, Maïa. 2018. A preliminary typology of emotional connotations in morphological diminutives and augmentatives. *Studies in language* 42(1) [Special issue: *Morphology and emotions across the world's languages*]. 17–50.

Rastorgueva, Valentina S. & Kerimova, Aza Alimovna & Mamedzade, Akhmed Kerimovich & Pireĭko, Lii͡a Aleksandrovna & Ėdel'man, Džoy I. 2012. *The Gilaki language* [English translation, editing and expanded content by Ronald M. Lockwood]. Uppsala: Acta Universitatis Upsaliensis.

Reinöhl, Uta. 2016. *Grammaticalization and the rise of configurationality in Indo-Aryan*. Oxford: Oxford University Press.

Reinöhl, Uta & Himmelmann, Nikolaus P. 2017. Renewal: A figure of speech or a process sui generis? *Language* 93(2). 381–413.

Roberts, Sarah & Bresnan, Joan. 2008. Retained inflectional morphology in pidgins: A typological study. *Linguistic typology* 12(2). 269–302.

Skjærvø, Prods Oktor. 1983. Case in inscriptional Middle Persian, inscriptional Parthian, and the Pahlavi Psalter. *Studia iranica* 12(1–2). 47–62, 151–181.

Stilo, Donald. 2004. *Vafsi folk tales*. Wiesbaden: Reichert.

Stilo, Donald. 2009. Case in Iranian: From reduction and loss to innovation and renewal. In Malchukov, A. & Spencer, A. (eds.), *The Oxford handbook of case*, 700–715. Oxford: Oxford University Press.

Stilo, Donald. 2016. Dikin Marāqei of Alamut: An undocumented conservative Tati language. Paper for presentation at the *Second International Symposium on Endangered Iranian Languages*, 8–9 July, 2016, CNRS, Paris.

Stilo, Donald. 2018. The Caspian region and South Azerbaijan: Caspian and Tatic. In Haig, G. & Khan, G. (eds.), *The languages and linguistics of western Asia: An areal perspective*, 659–829. Berlin: Mouton de Gruyter.

Viti, Carlotta. 2015. On degrammaticalization: Controversial points and possible explanations. *Folia linguistica* 49(2). 381–419.

Willis, David. 2016. Exaptation and degrammaticalization within an acquisition-based model of abductive reanalysis. In Norde, M. & Van de Velde, F. (eds.), *Exaptation in language change*, 197–225. Amsterdam: Benjamins.

Windfuhr, Gernot. 1992. Cases. In *Encyclopaedia iranica*, vol. V.1, 25–37. California: Mazda. (`http://www.iranicaonline.org/articles/cases`) (Accessed 2019-07-31).

Yarshater, Ehsan. 2009. Four Tati sub-dialects. In Sundermann, W. & Hintze, A. & de Blois, F. (eds.), *Exegisti monumenta: Festschrift in honour of Nicholas Sims-Williams*, 551–570. Wiesbaden: Harrassowitz.

Ylikoski, Jussi. 2016. Degrammaticalization in North Saami: Development of adpositions, adverbs and a free lexical noun from derivational and inflectional suffixes. In Hasselblatt, C. & Wagner-Nagy, B. (eds.), *Finnisch-Ugrische Mitteilungen*, vol. 40, 113–173. Hamburg: Buske.

Zahedi, Keivan & Mehrazmay, Roghayeh. 2011. Definiteness in Sorani Kurdish and English. *Dialectologica* 7. 129–157.

7
Language choice and patterns of usage among Kurdish speakers of Duhok: An empirical intergenerational study

Geoffrey Haig & Baydaa Mustafa

1 Introduction

It is well-known that the historical Kurdish speaking region is divided between four countries, Turkey, Iran, Iraq, and Syria, each of which has pursued different policies on Kurdish, ranging from prohibition of the language to various degrees of tolerance (Sheyholislami 2015). Outside of the Kurdistan Region of Iraq, Kurdish generally has little or no official status, and correspondingly lacks institutional support. The causes and consequences of this state of affairs have been extensively discussed in the literature, with reference to political ideologies, to education policies, and to international linguistic human rights. Notably, most of the readily accessible literature deals with the Kurds of Turkey (Çağlayan 2014; Haig 2004; Haig & Öpengin 2014; Öpengin 2012; Skutnabb-Kangas & Bucak 1995, among many others). For the Kurdistan Region of Iraq, however, where Kurdish has been an officially-recognized language in education and the public sphere for many decades, there is surprisingly little reliable research available on even the most basic issues of language choice and language attitudes.[1]

[1]For example, in 2012, a special edition of the *International Journal of the Sociology of Language* was dedicated to Kurdish, but contained no contribution on Iraq (Sheyholislami et al. 2012).

The present study targets language choice and language attitudes among Kurdish speakers in the multi-lingual city of Duhok[2] (Kurdistan Region of Iraq). While the main language of wider communication in Duhok city is the Bahdini dialect of Kurmanji, several languages (Bahdini, Sorani, Arabic, and English) have been used as the language of instruction in education during different periods of time, yielding an age-graded, multi-lingual community. Within such a community, different languages are selected for different settings (with different interlocutors, and in different contexts), and exercising language choice seems to be a natural, automatic and unplanned process, with obvious parallels to the choice of an appropriate register, genre, style, medium, or tone of voice in any communicative setting (Dweik & Qawar 2015). To date, no empirical sociolinguistic research of this nature has been conducted in Duhok, or indeed in any urban center of the Kurdistan Region of Iraq. In this chapter, we present the first results of an ongoing research project that focuses on the sociolinguistic variable of age, across a variety of attitudinal and usage-based parameters.[3] We have opted to look at age, because age has been generally less researched than other social variables such as gender, ethnicity or social class, and because age is crucial to understanding the link between language variation and change (Llamas 2007; Milroy & Gordon 2003). As Llamas (2007: 69) puts it:

> The treatment of age in sociolinguistic studies is influenced, to a degree, by a primary concern with language change or with language variation. Variationist, quantitative studies investigating language change in progress may approach chronological age as a methodological device with which to group speakers and to measure sociolinguistic differences across age groups.

Our preliminary results indicate that certain aspects of language usage and attitudes do correlate with age, though it is mainly the oldest age cohort (over 50) which differs significantly from the rest, and additional effects of gender are also apparent. We present our initial results with a minimum of descriptive statistics at this stage, and will restrict ourselves to identifying what appears to be some major trends, while deferring more complex analysis to the later stages of this project.

[2] We have adopted this spelling of *Duhok* in accordance with the official usage of the Duhok municipal authorities.

[3] We are extremely grateful to the many people of Duhok that participated in the interviews, and to an anonymous reviewer for extensive comments on earlier versions of this paper. We of course bear the sole responsibility for the remaining shortcomings.

In Section 2, we briefly outline the current language situation in Duhok city, and in Section 3 we summarize the main developments with regard to language of instruction in schools of the region. Section 4 outlines the project background and data sources and methodology, while Section 5 presents a selection of quantitative findings that have emerged so far. In Section 6, we close with a prospective outlook for future research in this direction.

2 Bahdini dialect in Duhok City

Within the Kurdistan region of Iraq, two varieties of Kurdish are spoken, Sorani and Bahdini. Sorani, also called 'Central Kurdish', is spoken by the majority of the Kurds, while Bahdini is spoken by around one million speakers. Linguistically, Bahdini belongs to the southeastern dialect group of Kurmanji (Northern Kurdish), according to the classification of Öpengin & Haig (2014).

While Sorani has an established written standard and is well represented in education and the media the status of Bahdini is complex and has seldom been considered in the literature. The following factors conspire to render Bahdini sociolinguistically disadvantaged in various ways:

- Within the context of the Iraqi state, and of international relations to other states, all varieties of Kurdish are disadvantaged in comparison to Arabic.

- Within the context of the Kurdistan Region of Iraq, Bahdini is the less-prestigious variety of Kurdish, when compared to Sorani.

- Within the context of Kurmanji (to which Bahdini belongs): Bahdini differs from the widely-accepted "standard variety" of Kurmanji, as codified in the grammar of Bedir-Khan & Lescot (1970), through a number of lexical and morphological features, which inhibit mutual intelligibility with speakers from the northerly dialects of Kurmanji, and have yet to be reliably analyzed. Thus, it is stigmatized as non-standard, dialectally divergent within Kurmanji itself.

- The vernacular of Bahdini is heavily overlaid with Arabic loans, further heightening the perceived distance to more-widely used varieties of Northern Kurdish, and adding to the stigmatization as an "impure" variety of Kurdish.

- Bahdini is written with the Arabic script, which inhibits participation in the trans-national, pan-Kurmanji cultural space, carried by the Roman alphabet-based Bedir-Khan & Lescot standard. The Arabic-based standard for Bahdini has not attained the same range of acceptance, or of standardization, as the Roman alphabet norm.

Despite the modest overall status of Bahdini, it remains the main language of everyday communication for most of the Kurds living in the traditional Bahdini regions. Our research focuses on Duhok city, one of the main centres for Bahdini, located in Duhok Governorate, in the north-west of Iraq. Duhok governorate forms the western governorate in the Kurdistan Region of Iraq and has a strategic location at the intersection of Syria, Turkey, and Iraq (Tovi & Badi 2010). Duhok Govenorate has an area of about 9755 km^2, and an estimated number of inhabitants of more than one million (Tovi & Badi 2010). The majority of the people are Sunni Muslims, but there are also a large number of Christians of different denominations, in addition to many Êzidi people. Duhok governorate is divided into seven districts: Duhok, Zakho, Amedi, Semel, Akre, Shixan and Bardarash (Tovi & Badi 2010).

3 Education system in Duhok City

With the establishment of the Iraqi state after the First World War, Kurdish people made demands to use their mother tongue as a medium of instruction in education. In 1930, the Iraqi parliament drafted new legislation to create the "Local Languages Law", in which linguistic minorities, such as Kurds, were granted some linguistic rights in their region (Sheyholislami et al. 2012). However, the primary goal of mother tongue education in Kurdish was not achieved until the late 1950's, with the exact extent and nature of implementation varying according to sources. In the 1960's, education policy shifted again to Arabic, but in the 1970's, again a brief period of education in Kurdish (Sorani) followed within a framework of an autonomy programme for the Kurdish region. However, Kurdish was used in schools for only five years, and afterwards all Kurdish schools were abolished and replaced by Arabic schools.[4] This continued until 1991 when Kurdistan proclaimed its autonomy.

In 1992, following the establishment of the Kurdistan region of Iraq, Kurdish was made the official language of the region. However, until the mid-

[4]Tovi, M. (June 5, 2016, personal communication).

1990's, standard Kurdish in Iraq, including Duhok city, meant Sorani Kurdish. This means that for the Kurdish people in Duhok City, they were still using a language in education which was not their mother tongue, and which for many children is, initially at least, largely incomprehensible (Haig 2007). Attempts were made to introduce Bahdini, the mother tongue of people in Duhok, into the curriculum in 1996 when the Duhok Assembly and Board of Education requested that the Kurdistan Regional Government (KRG) in Erbil assist them in implementing Bahdini instead of Sorani in schools, starting with grades one to three. By 2002, Bahdini was the medium of instruction for grades one to six. By 2012, the Board of Education in Duhok completely replaced Sorani with Bahdini in all the school grades (Sheyholislami 2015). However, uncertainty continues to prevail, and since 2003, many private schools have been opened in Duhok, where the language of instruction is English. English is in many ways an obvious choice, as it sidetracks some of the political issues that are inevitably associated with the main local languages (Arabic, Sorani Kurdish, Bahdini Kurdish), though of course English is far from being politically or ideologically 'neutral' (recall Iraq's colonial heritage under the British mandate in the first half of the 20th century, and the association of English with the American military presence in Iraq). In the 2015–2016 academic year, the Kurdistan Region of Iraq began a trial phase to change the language of instruction from Kurdish to English, but restricted it to the subjects of mathematics and the natural sciences.[5] Several schools have been selected from each city to implement the new system, and courses have opened for teachers to learn the new curriculum and English language. From the academic year 2016–2017, it is intended to apply the plan to all schools in Kurdistan region of Iraq. From 2011 onwards, some departments in different public universities (e.g. Soran University and Duhok University) have started to use English as the sole language of instruction in a number of degree programmes, such as sociology and political science (Sheyholislami 2015). In a new communiqué posted on their website, the Ministry of Higher Education in the Kurdistan Region of Iraq underscores the importance of English and outlines how the ministry in the last two years has been strengthening its efforts for English as a second language and the language of science.

[5]Suleiman, W. (June 7, 2016, personal communication).

4 Background and methodology of the study

After this brief overview of language choice and status in the education system of Duhok, we present some findings from an ongoing investigation into language attitudes and language usage among Bahdini speakers in Duhok and its environs (Mustafa In prep). Specifically, this paper addresses the following questions:

1. Do differences in age correlate with different levels in the degree of Arabic words used in the lexicon?

2. Do age differences correlate with attitudes towards choice of different languages in education?

3. Do age differences correlate with differences in the extent to which the Kurdish language is considered important for Kurdish identity ('being Kurdish')?

The data stem from a survey carried out with 108 adult speakers of Bahdini Kurdish (see Table 2 in the Appendix for a breakdown of all participants across age and gender). As we were particularly interested in the respective effects of different languages in education, we split our sample into three age cohorts: 'Generation one', aged between 18–30, have had their whole school education in Kurdish (N=34, 16 males and 18 females). The second group, labeled here 'Generation two', consists of persons aged between 31–50 who had their schooling in Arabic with one subject in Kurdish (N=40, 22 males and 18 females). The third group, 'Generation three', includes speakers over fifty years of age who had Arabic as the language of education in school (N=34, 16 males and 18 females). The choice of these three age groups thus approximately reflects the major changes in language of education across the last 50 years. We deliberately excluded speakers under 18 to minimize possible maturation effects in the data. We are well aware of the inherent problems in any kind of age divisions among adults; in an overview of the relevant literature, Eckert (1997: 165) notes that "adulthood has emerged as a vast wasteland in the study of variation", reflecting a general lack of consensus on the impact of age on patterns of language usage among adults. As mentioned, our division into three groups is dictated by the hypothesis that changes in language of education may have affected language use. It is nevertheless notable that the three divisions adopted here correspond to the divisions of young adult,

middle-aged, and older speakers that are widely adopted (though seldom explicitly justified) in much sociolinguistic research.

While the sample is reasonably balanced across age and gender, under the prevalent fieldwork conditions it was not possible to achieve a balanced sample across other social variables, such as socioeconomic class, level of education, religious and tribal (*aširet*) affiliation (for example, among the respondents, there are 13 members of the Êzidi community, 6 males and 7 females), as this would have involved screening a much larger pool of potential respondents in order to obtain sufficient respondents. In the metadata obtained from each participant, however, extensive additional information on speaker background has been systematically recorded, so the impact of these factors can be controlled for in later analyses. In the present context, we restrict ourselves to investigating the main independent variables of age and gender.

The methodology involved a three-part sociolinguistic interview, comprising of (i) a free speech section, (ii) a picture-naming task, and (iii) a questionnaire with sections on language choice in different communicative domains, and media consumption. The interviews were conducted in Duhok and the neighbouring townships of Sharya, Akre, Zakho and Bamerne by a female native speaker of the region. The content and methodology for the three parts is as follows:

1. In the free speech section, respondents were asked to describe the last wedding celebration they had attended, and to comment on wedding celebrations and how they had changed in Duhok in general. Additional questions were prepared (e.g. regarding food, dance, dress customs, etc.) in the event of participants' ceasing their accounts too quickly, but in most cases the topic proved highly suitable, and participants were readily able to talk freely for about 10 minutes.

2. The picture-naming task was based on a set of 42 pictures, of which 32 were target items while the remaining 10 were distractors. All the pictures were colour-printed and bound into a book format, so that the interviewer simply turned the pages, and the respondents named what they saw. The pictures represent objects that typically triggered a lexical choice between an Arabic vs. a Kurdish word, and which were deemed sufficiently familiar to all speakers.

3. A questionnaire consisting of four main parts: The first part contains questions regarding language choice in the media, the second concerns

language choice according to interlocutor, the third part is about language choice in social domains, and the last part concerns language attitudes.

All interviews were conducted in colloquial Bahdini with no special attempt to avoid Arabic elements on the part of the interviewer, and recorded from start to finish in a non-compressed data format (WAV). Questions were administered orally and answers entered into the forms by the interviewer, because many of the respondents cannot read or write. A total of 108 interviews were conducted although some data had to be excluded from certain analyses, for example when a picture was not properly recognized, or when a question was misunderstood.

The data were collected through a snowball sampling method, based on the interviewer's own social network. This methodology has obvious advantages and disadvantages. From a conceptual perspective, it is somewhat problematic because the sample is not randomly selected. However, from the perspective of fieldwork conditions in Iraqi Kurdistan, where familiarity and trust are crucial to enabling interviews to be conducted in a domestic setting, it is probably the only practicable method for reaching an adequate number of participants within a reasonable time-frame (see Milroy 1987, cited in Rasinger 2013, for discussion).

5 Research questions and results

5.1 Use of Kurdish in the picture-naming task

The first question concerns the correlation between age and choice of Kurdish or Arabic, when there are two words available for one meaning. We hypothesize that the older speakers, whose education was primarily through the medium of Arabic, will have higher levels of Arabic and lower levels of Kurdish in their lexical choices than younger speakers who have undergone education through the medium of Kurdish. We quantified "use of Kurdish" by calculating the number of Kurdish words used in the picture-naming task.

The picture-naming task minimized the verbal input of the interviewer, avoided the problems of a translation-based stimulus, reduced possible accommodation to the interviewer, and yielded a rich structured data set in a short time (about one minute with some interviewees); we refer to, e.g., Schmid & Köpke (2009) on picture-naming tasks as a data source for assessing lexical knowledge. Typical examples of the pictures that were used include

a picture of a restaurant, which could be named either *met'em* (Arabic) or *xwaringeh* (Kurdish). The Kurdish version of this word is quite transparent and has been used for many years (since the beginning of the establishment of the Kurdish Region in Iraq) in media and education, and is regularly written on the relevant signs outside restaurants. However, both the Arabic and the Kurdish word remain in use in everyday life. Another example is the picture of the flag, which could be named either *'elem* (Arabic) or *ala* (Kurdish). The Kurdish version of this word has also been used for many years in education and media. It is used each Thursday in schools when the flag is raised, but again, in everyday usage, the Arabic and the Kurdish words continue to coexist. Another example is 'teacher', which could be *mua'lim* (Arabic) and *mamosta* (Kurdish). Note that folk perceptions of etymology are not necessarily in line with philological facts. Thus for the meaning 'street' the Arabic word is widely used (*şari'*), but more recently, *cade* has been introduced through Kurdish-language media, and is assumed by most people to be a word of Kurdish origin. However, etymologically it is actually also of Arabic origin. For our investigation, we have taken common perception as the criterion for word origin. In most cases, the member of the word pair that we designate as 'Kurdish' is one that has been more recently introduced, primarily through Kurdish-language media and education. Our investigation monitors the degree to which speakers of different age groups recall and use these words.

In the current context, the methodology worked smoothly with the first and the second generations, but with the older generation, some unforeseen difficulties arose. Some of the elderly people had difficulties with the task due to poor eyesight and could not properly recognize the pictures, while others needed additional explanation to help them name the pictures. Furthermore, not all the stimulus pictures proved equally suitable. For some, respondents were unsure of the intended item, or named the wrong part of the picture. To minimize these effects, we excluded all items from the stimulus set which yielded this kind of ambiguous or uninformative response more than four times. This left 20 items on which the following analysis is based. In addition, we excluded the responses of two informants (two women, 83 and 71 years old), because most of their answers were not interpretable due to eyesight or other difficulties.

The percentages of Arabic, Kurdish, and mixed responses (i.e. respondents supplied two words, one from each language), distinguished according to generation, are shown in Figure 1; see Table 3 in the Appendix for the absolute figures.

In order to compare these groups, we used the *t*-test for independent samples, taking the respective means from each group in the value 'Kurdish', and compared them pair-wise. The results are provided in Table 1.

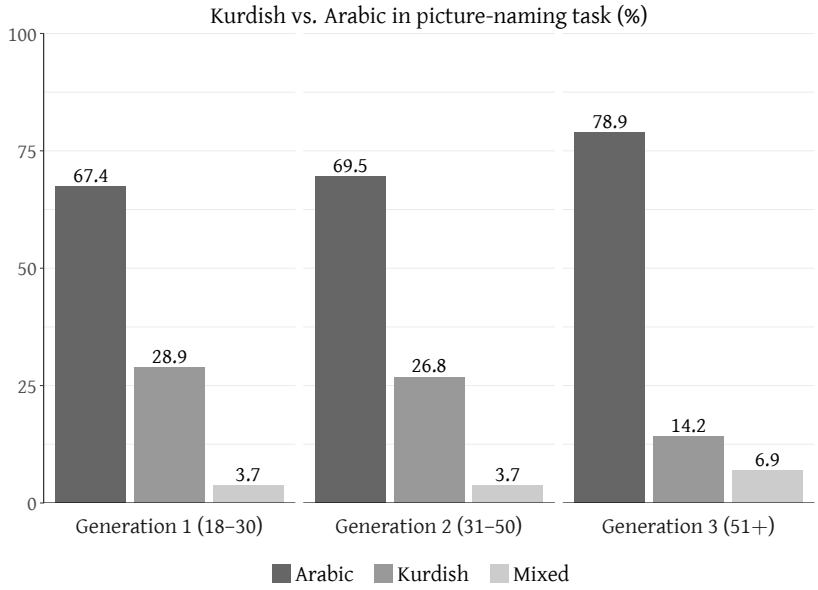

Figure 1: Percentage of Kurdish and Arabic responses in a picture-naming task, by age.

Table 1: Comparison of group means (choice of 'Kurdish') in the lexical decision task, using the *t*-test

Generations	*t*-value	*p*-value	Significance
Gen. 1 and Gen. 2	0.48216	0.315576	*not* significant at $p < 0.05$
Gen. 2 and Gen. 3	2.69367	0.004419	significant at $p < 0.05$
Gen. 1 and Gen. 3	2.97019	0.002093	significant at $p < 0.05$

According to the *t*-test, the difference between Generations 1 and 2 is not significant, but the difference between the oldest generation (Generation 3) and the other two is highly significant ($p = 0.004419$ and 0.002093, respec-

tively). This indicates that there has been a shift towards an increased use of Kurdish – at least under the conditions of our investigation – among adults of fifty years and younger. Breaking down the results according to respondent gender yields the figures given in Figure 2.

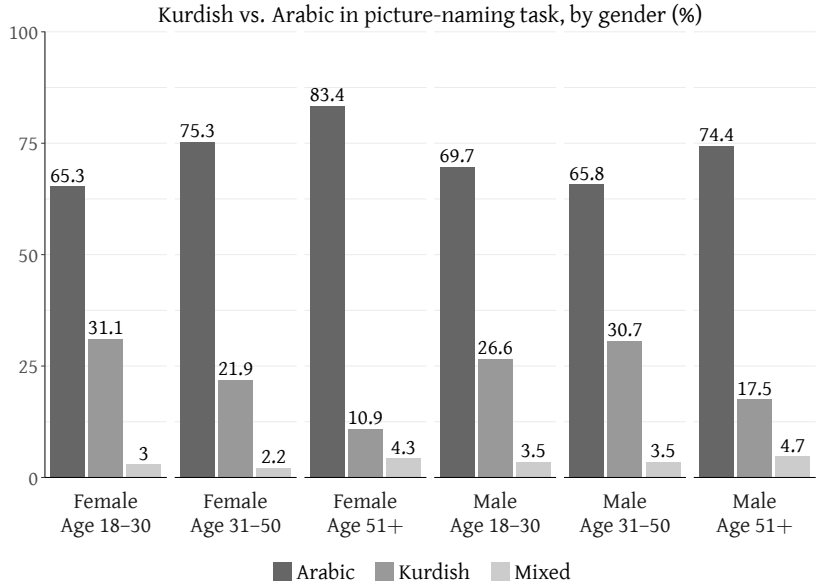

Figure 2: Percentage of Kurdish and Arabic responses in a picture-naming task, by age and gender.

Although the mean values shown in Figure 2 are suggestive of gender-specific differences within each generation, a *t*-test comparing the two genders within each generation yields no significant differences in the numbers of Kurdish items chosen. This is in part due to the small absolute values for each group (between 16 and 22, see Table 2, Appendix), which approach the widely-assumed lowest limit of 10 values per sample for applying the *t*-test.

Discussion: Considering first only generational differences (cf. Figure 1), there is a significant difference between the oldest generation and the two younger generations, with the older generation using an overall lower number of Kurdish words in their lexical choices. This is in line with the initial

hypothesis that those who underwent socialization and early education in an Arabic-based system will retain higher rates of Arabic in their lexicon, while younger speakers exposed to Kurdish language education will have adopted more Kurdish words.

With regard to gender effects, we do not find significant intra-generational differences, though we note that this may also be an artifact of the small sizes of the compared groups. Given that Arabic and Kurdish differ in terms of status, one might have expected greater gender-based differences. It has often been noted that women avoid "stigmatized variants" (Labov 2001: 266) in their speech to a greater degree than men do (Trudgill 1972, see Labov 2001: 263–272 for discussion). While Labov's claim is based on men and women's use of linguistic variables in a monolingual context (e.g. the choice between *-ing* and *-in* on English verb forms), the basic principle is also applicable to language choice. In other words, we would expect that if there is a prestige imbalance among the languages used in a multi-lingual setting, women will be statistically more likely to choose the higher-prestige language than men. Gal's (1979) case study of language choice among Hungarian/German bilinguals in Austria reveals that women are more likely to prefer German (associated with higher external prestige) than Hungarian, though the difference is only significant among the youngest generation (Gal 1979: Ch. 6). Similarly, Çağlayan (2014) underscores the role of women as leading the shift towards Turkish in Kurdish/Turkish bilingual families in Diyarbakir.

However, the Duhok example cannot be directly compared to the latter two case studies. With regard to the choice between German and Hungarian in the context of contemporary rural Austria, or between Turkish and Kurdish in Turkey, there is no doubt which language has the higher prestige in terms of social mobility, professional advancement, and economic opportunity. In the Duhok case, however, we witness a much more complex setting, where Arabic is the language of wider communication and cultural prestige in the context of the Iraqi state, and the broader Islamic cultural sphere, yet Kurdish enjoys local prestige, and since the establishment of the Kurdistan Region of Iraq, has been intensely promoted by the regional government, and linked to the cause of Kurdish nationalism. The picture is thus considerably more complex, and is unlikely to be accountable in terms of a mono-causal effect of gender-related differences in response to external prestige. The lack of any clear directionality across the genders that we find in our data may thus reflect the lack of (or shifting nature of) a prestige asymmetry among the languages concerned, but this requires confirmation over a larger sample, which is beyond the scope of this paper.

5.2 Language choice in education

The second part of our study here concerns speakers' preferences for the language of instruction in schools. Among the responses, three options were overwhelmingly preferred: English, Bahdini, and English together with Bahdini; the raw figures are provided in Table 4 in the Appendix, and only distinguish these three options.[6] We consider the responses to this question to be an important indicator of language attitudes, and probably a more reliable measure of genuine attitudes than questions that overtly target such attitudes (e.g. 'Do you have a positive attitude to language X?', or similar). Of course this question also involves self-reporting on a highly politicized issue, and is thus not without its drawbacks, and in a more convenient fieldwork setting, other methodologies such as matched-guise techniques would have been preferable. But given the constraints of the interview setting, the range of options for obtaining reliable information on this topic was fairly limited. Despite the methodological challenges, this question turned out to yield interesting and not obviously predictable results. Note that only a few informants entered "Arabic" as a preference (cf. Footnote 6), indicating the general rejection of Arabic-based education, presumably reflecting the continued perception of Arabic as a symbol of political and cultural oppression, although other reasons may also be relevant.

The Fisher's Exact Test of the difference between Generations 1 and 2 with regard to choice of English, and choice of Bahdini, reveals a Fisher exact value of 1, which is not significant at $p < 0.05$. The difference between Generations 2 and 3, however, is highly significant (Fisher exact value 0.006484). Figure 4 gives the results according to gender differences in each group. Due to the small numbers in each group, we have not conducted significance testing.

Discussion: Considering Figure 3, it is evident that the main generational difference lies between Generations 2 and 3. In this case, there is a striking increase in preference for English, either as the sole medium of education, or in combination with Bahdini. While around 23% of the oldest generation chose either English, or English and Bahdini, this figure rises to over 67% in the youngest respondents. Interestingly, the preference for English-only is more pronounced in Generation two than in Generation 1. There is also

[6] Among the marginal options which were excluded were participants who chose Arabic and English, or Arabic with English and Bahdini, and another one who chose Latin. None entered Sorani as a preference.

a slight, but consistent, difference between the genders with regard to attitudes towards English. In all generations, English is a more popular choice for women than for men. Although the differences within each generation do not reach statistical significance, it is notable that in each generation, the difference is in the same direction. Whether this can be interpreted in terms of the tendency for women to prefer languages associated with social mobility, as discussed above, remains an open question.

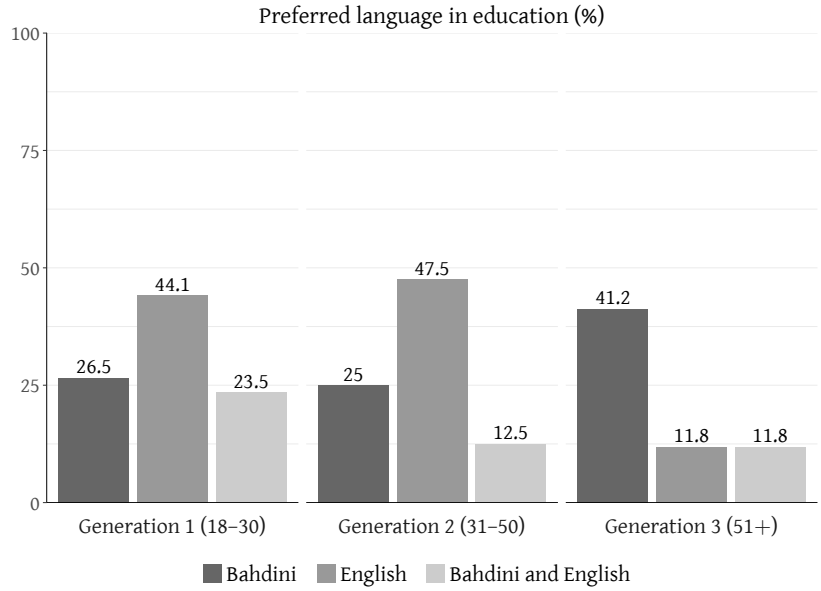

Figure 3: Preferred language in education, by age.

The increased preference for English among the younger two generations is probably a result of the higher proportion of university-educated respondents, who are familiar with the prestige of English as the international language of science and internet-based communication, hence a vehicle for upward mobility and social advancement. The Kurdistan Regional Government has provided many scholarships for students working in different fields to do their higher education abroad and has heightened the perception of English as the key career choice in academic and professional settings (Barbarani 2013). It may also reflect disappointment with the implementation

of Kurdish language education in the education system, with English providing a compromise solution, still preferable to the ideologically stigmatized Arabic option, but avoiding the practicability issues that Kurdish language education faces. There is a growing demand for (and availability of) private schools in which the language of instruction is English (Barbarani 2013). In the youngest generation, it is noteworthy that Bahdini is still favoured by around 37% of the male respondents, while only around 16% of women choose this option.

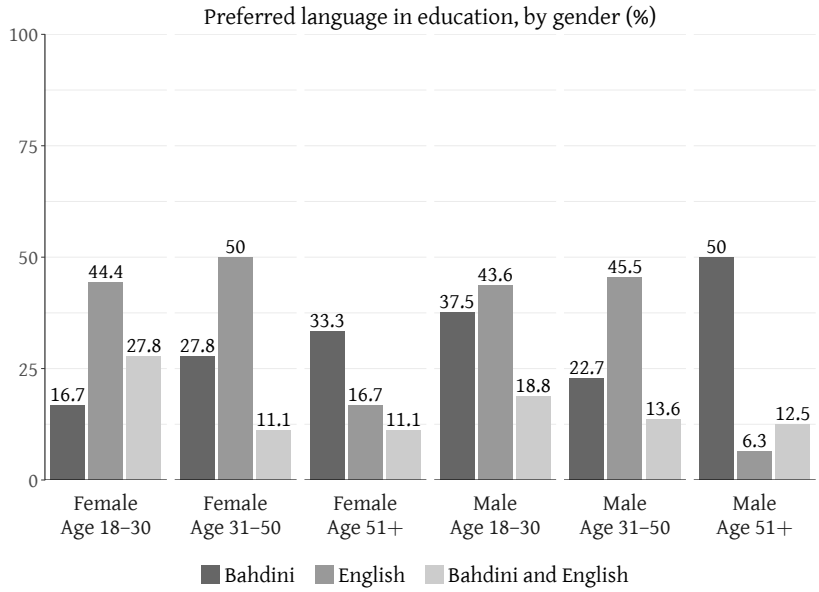

Figure 4: Preferred language in education, by age and gender.

5.3 Is Kurdish important for Kurdishness?

The final data we look at here concern the role of the Kurdish language for Kurdish identity. In the questionnaire, we asked the question in terms of 'whether it is important to speak Kurdish in order to be a Kurd', because dealing with abstract academic concepts such as 'identity' is not practicable in this fieldwork situation. This question is intended to complement the pre-

ceding one, which refers to language choice in education. Figure 5 gives the respective percentage of yes/no answers to this question for each generation, while Figure 6, provides the same information broken down for gender.

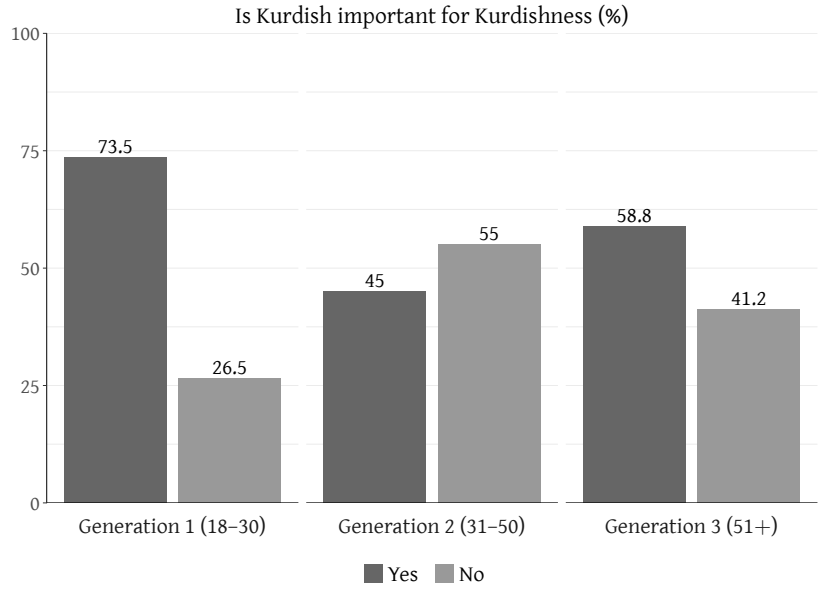

Figure 5: Responses to the question *Is Kurdish important for Kurdishness?,* by age.

With regard to generational differences, a Fisher's Exact Test yields a very significant difference between generations one and two (Fisher exact value 0.018227), while the difference between Generations two and three does not reach significance (Fisher exact value 0.253964). The difference between the two endpoint Generations one and three also does not reach significance (Fisher exact value 0.21424). We have not tested the intra-generational gender differences (Figure 6) for significance due to the low absolute values in some of the cells (cf. Appendix, Table 5).

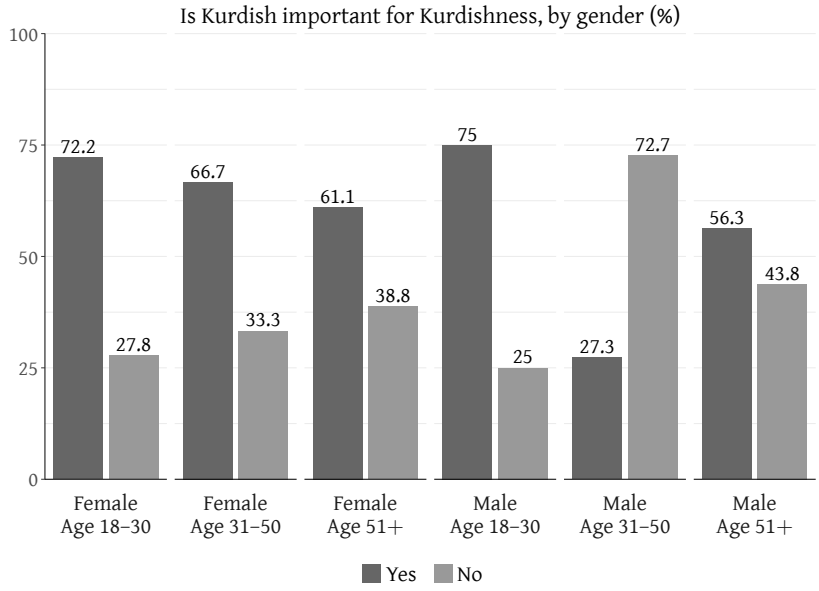

Figure 6: Responses to the question *Is Kurdish important for Kurdishness?,* by age and gender.

Discussion: In a simple world, we might have expected a correlation between the importance attached to knowledge of Kurdish for 'being Kurdish' (Figure 5), and choice of Kurdish as a medium of education (Figure 3). The results do not confirm this expectation. Figure 5 shows that the younger generation attach significantly more importance to speaking Kurdish than Generation two. Yet in Figure 3 above, we see that the younger generation is significantly less in favour of education solely in Bahdini than the older generation. For the youngest generation, then, active command of Kurdish is apparently linked to a notion of Kurdishness (approximately three quarters of the respondents in this group answered the question with 'yes'), but this belief is not matched by a desire to promote Kurdish as a language of education.

The data also reveal a striking effect of gender in Generation 2. In the other age cohorts, a majority considers that knowledge of Kurdish is important for being Kurdish, but in Generation two, fewer people overall share this view, and the breakdown of the group according to gender (Figure 6) reveals that it is the men of this age group that overwhelmingly responded with 'no'. Note also that the same group (males 31–50) show a general dispreference for Kurdish in education, with only some 23% of respondents choosing this option (cf. Figure 4), the lowest among any of the male groups in our sample. We can only speculate on the reasons for the lack of importance attached to the Kurdish language among members of this group. A glance at Table 2 in the Appendix reveals that most of the males recruited to this cohort are in their thirties, i.e. born in the 1980's, and would have experienced the traumatic and violent period around the transition to autonomy in their childhood, and the early days of autonomy as young adults. It is possible that this may have negatively impacted on their attitudes towards the Kurdish language, but this awaits a more detailed study of this group, with a larger sample. However, piloting this question in our investigation has unearthed a potentially very significant age and gender effect with regard to attitudes towards language and identity.

The responses of the younger generation, on the other hand, seem to reflect a fairly solid association of Kurdish language with Kurdish identity, stable across both genders. Whether this reflects a genuine conviction among the respondents, perhaps interpretable as a degree of success for the promotion of the Kurdish language in the Kurdistan Region of Iraq, or simply growing awareness of the issue of identity among the younger generation, and a desire to present themselves as 'pro-Kurdish' in the interview situation, cannot be answered with certainty. Regardless of the causes, however, there is clearly a very significant shift in behaviour with respect to this issue that marks the younger generation from the two older ones.

6 Conclusion

This pilot study investigates the relationship of language use and language attitudes among Kurdish speakers in Duhok, focusing on age-based differences. We have only considered the impact of two independent variables at this point, age and gender, but even this has revealed a complex picture that invites more detailed investigation. Nevertheless, certain trends emerged as fairly robust, and may serve as an anchor for future research. First, we were

able to show that the generation of over-fifty-year-old speakers use significantly fewer Kurdish words in a lexical decision task than younger speakers. Whether this result reflects genuine usage, or performance in a monitored setting (where younger speakers may have been consciously avoiding non-Kurdish items) is difficult to ascertain. Whichever explanation (or combination thereof) is ultimately responsible, we can nevertheless state with some confidence that age does indeed impact on linguistic behaviour. With regard to language attitudes, we also found age effects, though to some extent in contradictory directions: on the one hand, speakers from the youngest generation are significantly more likely to consider knowledge of Kurdish to be important for 'being Kurdish' than the oldest generation (Section 5.3). On the other hand, the youngest generation actually expresses less support for Kurdish as a medium of education than the oldest generation (Section 5.2). We tentatively interpret this in terms of the practical difficulties that have been experienced by the younger generations (and their children) in the nascent Kurdish language education system.

Suggestive evidence of gender effects have also been found, though they do not reach significance, in part due to the low absolute figures involved when age cohorts are split according to gender. A gender effect that was consistent across all three generations was a higher preference for English as a medium of education among women than men. This may reflect the tendency noted in other studies (Gal 1979) that women are more likely to choose languages that offer greater prestige than men are, but this would require a larger sample in order to be verified, though more research is required to verify these effects, and to address their underlying causes.

Our research indicates cross-generational differences, both in language usage (levels of Arabic in the lexicon) and in attitudes. The ongoing analysis of other person-related factors (media consumption, language usage according to domains) will yield a more complete picture of what is evidently a very dynamic linguistic ecology. Additionally, the linguistic variables that can be identified in the free speech sections of the interview (not analysed here) will add a further layer to our understanding of inter-generational language change in Duhok. We hope that our work will stimulate further research on the multilingual context of the Kurdistan Region of Iraq, and that the evidence-based approach reported here may inform future policy-making in the field of language choice in education.

References

Barbarani, Sofia. 2013. The rise of English in Iraq's Kurdistan region. *Rudaw*. (http://www.rudaw.net/english/kurdistan/29042013) (Accessed 2019-01-13).

Bedir-Khan, Emir Djeladet & Lescot, Roger. 1970. *Grammaire kurde (dialecte kurmandji)*. Paris: Maisonneuve.

Çağlayan, Handan. 2014. *Same home, different languages. Intergenerational language shift: Tendencies, limitations, opportunities. The case of Diyarbakır*. Diyarbakır: Diyarbakır Institute for Political & Social Research.

Dweik, Bader S. & Qawar, Hanadi A. 2015. Language choice and language attitudes in a multilingual Arab Canadian community: Quebec-Canada: A sociolinguistic study. *British journal of English linguistics* 3(1). 1–12.

Eckert, Penelope. 1997. Age as a sociolinguistic variable. In Coulmas, F. (ed.), *The handbook of sociolinguistics*, 151–167. Oxford: Blackwell.

Gal, Susan. 1979. *Language shift: Social determinants of linguistic change in bilingual Austria*. New York: Academic Press.

Haig, Geoffrey. 2004. The invisibilisation of Kurdish: The other side of language planning in Turkey. In Conermann, S. & Haig, G. (eds.), *Die Kurden: Studien zu ihrer Sprache, Geschichte und Kultur*, 121–150. Schenefeld: EB-Verlag.

Haig, Geoffrey. 2007. *A note on language politics in Iraqi Kurdistan and their repercussions for pan-Kurdish cultural unity*. (https://www.academia.edu/40081103/A_note_on_language_politics_in_Iraqi_Kurdistan_and_their_repercussions_for_pan-Kurdish_cultural_unity) (Accessed 2019-08-20).

Haig, Geoffrey & Öpengin, Ergin. 2014. Kurdish: A critical research overview. *Kurdish studies* 2(2) [Special issue: *Kurdish linguistics: Variation and change*]. 99–122.

Labov, William. 2001. *Principles of linguistic change*. Vol. 2: *Social factors*. Oxford: Wiley-Blackwell.

Llamas, Carmen. 2007. Age. In Llamas, C. & Mullany, L. & Stockwell, P. (eds.), *The Routledge companion to sociolinguistics*, 69–76. London: Routledge.

Milroy, Lesley. 1987. *Language and social networks*. Oxford: Blackwell.

Milroy, Lesley & Gordon, Matthew. 2003. *Sociolinguistics: Methods and interpretation*. Oxford: Blackwell Publishing.

Mustafa, Baydaa. In prep. *Language choice and language attitudes among Kurdish speakers of Dohuk: An empirical intergenerational study*. Bamberg: University of Bamberg dissertation.

Öpengin, Ergin. 2012. Sociolinguistic situation of Kurdish in Turkey: Sociopolitical factors and language use patterns. *International journal of the sociology of language* 2012(217). 151–180.

Öpengin, Ergin & Haig, Geoffrey. 2014. Regional variation in Kurmanji: A preliminary classification of dialects. *Kurdish studies* 2(2). 143–176.

Rasinger, Sebastian M. 2013. *Quantitative research in linguistics.* 2nd edn. London: Bloomsbury.

Schmid, Monika & Köpke, Barbara. 2009. L1 attrition and the mental lexicon. In Pavlenko, A. (ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (Bilingual Education & Bilingualism), 209–238. Bristol: Multilingual Matters.

Sheyholislami, Jaffer. 2015. Language varieties of the Kurds. In Taucher, W. & Vogl, M. & Webinger, P. (eds.), *The Kurds: History, religion, language, politics*, 30–51. Vienna: Austrian Federal Ministry of the Interior.

Sheyholislami, Jaffer & Hassanpour, Amir & Skutnabb-Kangas, Tove (eds.). 2012. *The Kurdish linguistic landscape: Vitality, linguicide and resistance.* [Special issue of the *International journal of the sociology of language* 217].

Skutnabb-Kangas, Tove & Bucak, Sertaç. 1995. Killing a mother tongue: How the Kurds are deprived of linguistic human rights. In Skutnabb-Kangas, T. & Phillipson, R. (eds.), *Linguistic human rights: Overcoming linguistic discrimination*, 347–370. Berlin: de Gruyter Mouton.

Tovi, M. & Badi, I. 2010. *Duhok governorate guide*. Duhok: Hawar Publishing.

Trudgill, Peter. 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in society* 1(2). 179–195.

Appendix

Table 2: Breakdown of all respondents across ages and genders

Generation 1			Generation 2			Generation 3		
Age	**N Male**	**N Fem.**	**Age**	**N Male**	**N Fem.**	**Age**	**N Male**	**N Fem.**
20	1	0	31	4	0	51	1	1
22	1	5	32	5	2	53	1	3
23	1	1	33	2	2	54	1	0
24	1	1	35	1	1	55	0	1
25	4	0	36	3	1	56	1	0
26	2	0	37	2	1	57	1	1
27	1	4	38	0	2	58	0	2
28	0	1	39	1	1	59	1	0
29	1	1	40	0	1	60	0	2
30	4	5	41	1	1	61	1	1
Sum	**16**	**18**	42	0	1	62	1	1
			43	0	1	63	0	1
			44	1	2	64	0	2
			46	1	0	67	1	0
			48	0	1	69	1	0
			49	1	0	70	1	1
			50	0	1	72	1	0
			Sum	**22**	**18**	73	2	0
						74	1	0
						75	0	1
						79	1	0
						82	0	1
						Sum	**16**	**18**

Table 3: Raw figures (group means) for Section 5.1 (language choice in the picture-naming task)

	Generation 1 (18–30)		**Generation 2 (31–50)**		**Generation 3 (> 50)**	
	Male (N=16)	Female (N=18)	Male (N=22)	Female (N=18)	Male (N=16)	Female (N=16)
Arabic	13.93	13.05	12.95	15.05	14.87	16.68
Kurdish	5.31	6.22	6.13	4.38	3.5	2.18
Mixed	0.56	0.55	0.77	0.38	0.75	0.68

Table 4: Raw figures for Section 5.2 (language choice in education)

	Generation 1 (18–30)		**Generation 2 (31–50)**		**Generation 3 (> 50)**	
	Male (N=16)	Female (N=18)	Male (N=22)	Female (N=18)	Male (N=16)	Female (N=16)
Bahdini	6	3	5	5	8	6
English	7	8	10	9	1	3
Bahdini & English	3	5	3	2	2	2

Table 5: Raw figures for Section 5.3 ('Is speaking Kurdish important for Kurdishness?')

	Generation 1 (18–30)		**Generation 2 (31–50)**		**Generation 3 (> 50)**	
	Male (N=16)	Female (N=18)	Male (N=22)	Female (N=18)	Male (N=16)	Female (N=16)
Yes	12	13	6	12	9	11
No	4	5	16	6	7	7

8
Temporal noun squishes in Kurmanji academic writing: From lexicality via NP-level junction to clausal subordination

Annette Herkenrath

Abstract: This is a synchronic investigation of the transitional area between clausal and NP-level patterns of junction, based on a corpus of academic writings published in Kurmanji Kurdish. By 'junction' is meant the linking of distinct syntactic units, ranging from NP-internal to clausal (i.e. by subordinators). A number of junctors in Kurmanji are multi-word units that combine a noun with adpositions, case, ezafe, deixis, indefinite determiners, phoric expressions, quantifiers, *wh*, plural, as well as the semantically neutral complementizer *ku*, enabling junction at a variety of syntactic levels. Academic writing in general can be assumed to purposefully shift between nominal and clausal patterns: achieving impersonal concision vs. providing slots for verb arguments. The study focuses on lexical nouns with a temporal meaning: *dem* 'time, period', *gav* 'moment, time, step', *wext* 'time, period, season' and *çax* 'time, age, period, era', which can flexibly change roles between lexical noun and subordinating junctor.

1 Introduction

This paper[1] takes a synchronic look at the transitional area between clausal and NP-level patterns of junction in a corpus of academic writings published in Kurmanji Kurdish. The term 'junction' refers to the linking of distinct syntactic units, both within the NP and at clause level (e.g. as subordinators). A number of junctors in Kurmanji are multi-word units that combine a noun with adpositions, case, ezafe, deixis, indefinite determiners, phoric expressions, quantifiers, *wh*, plural, as well as the semantically neutral complementizer *ku*, enabling connective employment at a variety of syntactic levels. Academic writing in general can be assumed to purposefully shift between syntactic levels, especially between nominal patterns (used to achieve concision and impersonality), and clausal ones (providing more slots for the distinctive expression of verb arguments). The present study focuses on a group of lexical nouns with a temporal meaning: *dem* 'time, period', *gav* 'moment, time, step', *wext* 'time, period, season' and *çax* 'time, age, period, era'.[2] These 'temporal nouns' ('TNs') can flexibly transcend categorial distinctions, to the effect that their syntactic role changes between that of a lexical noun and that of a subordinating junctor. Example (1) contains a junctor-like (1a) and a nominal (1b) employment of the TN *dem* 'time, period':

(1) *Di hevdîtin-an de yek wan xal-ên*[3] *ku herî*
CRP interview-OBL.PL CRP one DEI.OBL.PL situation-EZ.PL COMP very
zêde berçav bû ew bû ku ***dem-a***
frequently visible be.PST.3SG DEI.RCT be.PST.3SG COMP time-EZ.F
[mamoste-yan qal-a astengî û problem-ên xwendekar-ên
teacher-OBL.PL talk-EZ.F challenge and problem-EZ.PL pupil-EZ.PL
xwe di-kir-in], ji ***dem-ên*** *[xwe yên dibistan-ê] referans*
RFL ASP-do.PST-PL PRP time-EZ.PL RFL EZ.PL school-OBL.F reference

[1] Work on this paper grew out of an interdisciplinary cooperation at Justus Liebig University, Giessen, involving a discussion group on academic writing initiated by Mathilde Hennig. The corpus design received valuable hints from discussants at the *Sixth International Conference on Iranian Linguistics*, Tbilisi, June 23–26, 2015, notably Carina Jahani, Stephen Levinson, and Ergin Öpengin; the present design is somewhat of a compromise. Corpus-linguistic methods together with some emerging analytical ideas received further discussion at the *Third International Conference on Kurdish Linguistics*, Universiteit van Amsterdam, August 25–26, 2016. Additional thanks for critical remarks go to Geoffrey Haig, Ergin Öpengin and two anonymous reviewers.

[2] Related expressions such as *car* 'time', *dewr* 'turn, period', and *cerg* 'ring' were also observed. Since the data showed no employment as clause subordinators, they were excluded from the investigation.

di-da-n *û* *behs di-anî-n-e* *ser*
ASP-give.PST-PL and talk ASP-bring.PST-PL-DIR PRP
serhati-yên *xwe.*
experience-EZ.PL RFL

'One of the situations that could very frequently be observed in the interviews was that when teachers talked about the challenges and problems of their pupils, they would refer to their own schooldays and bring the topic to their own experiences.' (KA_025_kur_t_073)[4]

The TN-construction *dema mamosteyan qala astengî û problemên xwendekarên xwe dikirin* (1a) can be translated as a temporal subordinate clause 'when teachers talked about the challenges and problems of their pupils'; the noun-plus-ezafe form *dem-a* 'time-EZ.F' functions as a clause-level junctor. By contrast, in *mamosteyan* [...] *ji demên xwe yên dibistanê referans didan* 'teachers [...] would refer to their own schooldays' (1b), *dem* means 'days, period, time'; someone's *demên dibistanê* are this person's schooldays. In both instances, *dem* is connected with its – noun-modifying – syntactic environment via ezafe, irrespective of the difference in syntactic roles.

As will become apparent, the full range of use does not fit into a dichotomy of subordinator versus lexical noun, as Example (1) might suggest. The present study looks at how TN-based junctors function in context, addressing the following questions: (1) How are multi-word expressions built on one lexical core employed to achieve effects of junction in different syntactic environments? (2) How do temporal nouns fare individually on an assumed scale between nominal lexicality and clause-level junction? (3) How do TNs combine with other elements, such as ezafe, adpositions, and *ku*? The analysis will present two paths of 'categorial gradience' ('categorial squishes' in the sense of Ross 1972): between nominal and clausal modifications to bare temporal nouns, and between temporal nouns as bare nouns versus accompanied by nominal functional categories.

Section 2 outlines the theoretical interest of the study; Section 3 presents the data; Section 4 inventories TN-constructions according to their phrasal characteristics. Section 5 and Section 6 establish two paths of gradience: Section 5 orders bare TN-ezafe-constructions according to modifier types on a

[3]One reviewer wondered whether there might be a preposition missing: *yek **ji** wan halên ku ...* Example (1) has been double-checked: it corresponds to the original wording. The issue might deserve further scrutiny, however, not with the present thematic focus.

[4]Temporal noun expressions are given in boldface. Square brackets are used to mark formal attributes to TNs. The last figure in the code (e.g. _073) is a page number.

scale between fully nominal and fully clausal (Squish 1). Section 6 studies clausal constructions attributed to functionally specified TNs, i.e. adpositional and determiner phrases (Squish 2). Section 7 intersects the two transitions. While the main goal is a data-driven qualitative categorisation, a quantitative context is provided to appreciate tendencies of preferred usage in the corpus.

2 Theoretical interest

Grammars of Kurdish mention the investigated expressions in some of their uses as subordinators: Bedir-Khan & Lescot (1970: 269–273, 265) mention *dema ko* and *wexta ko* as subordinating conjunctions corresponding to French 'lorsque' and *gava* as 'lorsque' or 'quand'. They mention a principal link between conjunctions and prepositions in Kurmanji, the nominal character of *gav* 'instant', as well as some adverbial usages (*vê gavê* 'maintenant', *wê gavê* 'alors', *gavekê* 'une fois', *gavina* or *gavgavina* 'parfois, de temps en temps'). Kurdo (1991 [1984 [1973]]: 283) mentions *çaxê ku, dema ku, gava ku* and *wextê ku* under the term 'pevgırêkên wext û demê' (temporal conjunctions), in their use with indicative verbs. Thackston (2006: 20–22, 72–73) analyses *çaxê ku, dema (ku), gava (ku)* and *wexta (ku)*, all translated as 'when', as composed of a preposition plus (sometimes optional) *ku*, without, however, listing *çaxa, dema, gava* or *wexta* in his chapter on prepositions.[5]

The present study has been inspired by three theoretical frameworks: academic writing research, polylexical junctor research, and approaches of gradient (or 'fuzzy') grammar. Academic language, which requires concision and impersonality, has cross-linguistically been noted for its tendency towards nominal style (recently Hennig 2015a, 2015b, 2016). On the other hand, clausal constructions offer slots for syntactic argument positions that may not be available in the structure a noun can project. For expressive reasons, academic writing can therefore be expected to move between nominal and clausal constructions, offering material for a study of possible analogies (Szabolcsi 1990, 1994; Ágel 2013; Czicza 2015). The issue calls for a comparative approach to academic language. Thielmann (2009) compares English and German with respect to linguistic actions carried out in academic texts

[5] Examples (2) and (6) of the present paper might be cases in point, displaying prepositional uses of *dem* and *çax*, respectively.

(reader steering, connectivity, and naming).[6] Heller (2012) compares Italian and German with respect to functions of deictic, derivational and other means of reader orientation. While none of these comparative studies specifically focuses on issues of nominality versus clausality, a central idea is that the inner functional composition of individual connective expressions can best be teased out when looking at their use in a variety of contexts.

Investigations of academic Kurmanji are still rare (Herkenrath 2016b). However, Kurmanji has a typologically specific interest: its productive cross-categorial use of ezafe facilitates the nouny expression of complex ideas. As head-dependent constructions, ezafe constructions can in a sense neutralise the distinction between arguments and modifiers, expressing both by means of a single construction type.[7] Furthermore, both neutral and semantically specific connecting elements can be shown to display a syntactically diverse employment (Herkenrath 2015); specific lexical nouns such as *dem* 'time', *ber* 'front', *qend* 'degree, amount' can form sets of polylexical or multi-word junctors ('families of junctors') that can be employed at different syntactic levels (Herkenrath 2016a).

Looking at Kurmanji Kurdish, the present study draws on analytical ideas cross-linguistically developed with respect to the semantics and micro-level functionality of internally complex connective elements in German (Redder 1990, 2007; Rehbein 1995; Fabricius-Hansen 2007) and Turkish (Borsley & Kornfilt's 2000 'mixed extended categories'; Kornfilt & Whitman 2012a, 2012b on nominalisation; Herkenrath 2014), two contact languages of Kurmanji, as well as in wider typological perspectives (Mithun 1988). Greaves & Warren (2010) present multi-word constructions as a corpus-linguistic topic of research. Libert (2014) cross-linguistically discusses the potential of nouns, including some with a temporal meaning, to be functionalised for clause-combining purposes. Kirchner (2006) analyses phenomena of cross-linguistic 'reconstruction' in temporal subordination, where strategies based on TNs are used on both sides of a Turkic-Iranian divide: in Persian, Ottoman, Azeri, and Turkish. Matras (1997a, 1997b, 2002) discusses the distribution of linguistic elements that constitute various types of complementizers in Kurmanji, however, without considering temporal clause linking. The TNs presently under investigation (*dem, gav, wext* and *çax*) are one specific group of junctor-

[6] Ehlich (1992), at the basis of some of these comparative studies, highlights the specifically complex attention-organising functioning of deixis in academic ('scientific') texts, also dealing with issues of temporal deixis. While his study is based on German only, his analytical model holds potential for comparative undertakings.

[7] Thanks to Geoffrey Haig for this useful precision.

forming lexical expressions that, working in a transitional zone between NP-level and clause-level junction, achieve effects of temporal connectivity in Kurmanji.

In an explorative theoretical framework of 'fuzzy grammar'[8], Ross (1972: 316) proposes two categorial continua ('squishes') in English, namely one that ranges from verbs to nouns via different types of participles, adjectives, and adjectival nouns, and a second one ranging from *that*-clauses to lexical nouns via prepositionally subordinated clauses, embedded *wh*-constructions, different types of *-ing*-constructions, action nominals, and derived nominals (Ross 2004 [1973]: 351).[9] Corver & van Riemsdijk (2001) suggest to equally look for squishiness at the lexical/functional borderline; they discuss projections of functional categories and related syntactic movement in 'semi-lexical nouns'. What is important about fuzzy grammar research are its synchronic perspective and its focus on just one language.

This framework may be more widely contextualised in terms of 'emergent grammar' – or, for that matter, the extensive discussion of 'grammaticalisation' – , which makes reference to concepts of gradience (Hopper 1987, Hopper & Traugott 1993: 177–178, Himmelmann 1992; Lehmann 2015 [1982], 1988). Most of these approaches are interested in diachronic issues: they describe the direction of a development. Others, such as Sapir (1921) follow synchronic typological interests, paying attention to semantic shades of categorial gradience. The phenomena presently under investigation may be related to Sapir's (1921: 86f., 123f.) typologically-inspired idea of a word's diachronic "passage through a categorial continuum" (Lehmann 2015 [1982]: 5f.), with syntactic and semantic shades between 'concrete' and 'relational' concepts; while an underlying lexical concept may become 'latent', 'latent' does not mean 'lost'.

The present study takes semantically specific lexical units as a starting point, in order to follow their formal and functional variablity within the data. It combines polylexical junctor research with gradient grammar research in order to look at open-class phenomena of junctors. It follows a

[8] See Bolinger (2004 [1961]) for an early discussion of phenomena of categorial overlap and ambiguity; Aarts et al. (2004) for a historical discussion of the concept; Bresnan's (1997) 'extended heads' in a discussion of constraints on syntactic combinability in 'hybrid constructions'.

[9] The research paradigm of 'fuzzy grammar' can be criticised for not forming "a coherent grammar theory", as one reviewer points out. The issue is, however, programmatic: the analytic goal of 'fuzzy grammar' research precisely consists in identifying areas that resist integration into a coherent model. Sapir (1921: 94) in this connection speaks of a 'destructive analysis of the familiar'. The entire undertaking must be seen as preliminary.

synchronic interest, related specifically to academic writing, qualitatively focusing on just one language: Kurmanji Kurdish. How does the gradient character of TNs enable their categorially flexible employment? How does it enable playing with the boundaries between verbal/clausal and nominal style, as is functional in academic writing?

3 Corpus of the study

The corpus (Herkenrath, In prep.), is a collection of contemporary academic writings published in Kurmanji Kurdish, thematically pertaining to the humanities and social sciences: linguistics, literary criticism, social, historical and cultural studies, and psychology. Spanning at present some 4,000 printed pages, the corpus contains a variety of text types, including for the most part academic writing in a narrower sense, but also some texts in which academic authors address a wider public, as well as a few samples published as transcriptions from oral academic speech; see Table 1.

Since Kurmanji cannot be considered to have been the main language of anybody's academic socialisation, in order to create academic texts, authors combine native competence in Kurmanji with patterns and conventions of academic writing acquired in other languages (Matras' & Reershemius' 1991 on orthographic issues of standardisation). Mirroring this overall situation, the corpus contains different types of individually or collaboratively produced academic Kurmanji, next to text versions published in other languages.[10] Abstracting away from these genre distinctions, the present study deals with an exclusively Kurmanji subcorpus of 1,490 pages, of which 632 pages were manually tagged and concordanced.[11]

[10] In cataloguing texts, a coding system was applied to enable tracking of parallel versions: monolingually published Kurmanji original texts (kur_m), translations from other languages into Kurmanji (kur_t), and parallel publications (kur_p). In the references, translators [tr.] are listed as second authors in view of their crucial role in the production of the Kurmanji forms. The full bibliographical information is given in the Appendix.

[11] While a common measure of corpus size is number of words, the present corpus is a pilot corpus that still awaits digitisation. This is why its size is presently measured in pages.

Table 1: Data overview

Code	Short reference	Short title	Pages overall	Concordanced
KA_001_kur_m	Akin 2013c	Pêşgotin	5	5
KA_002_kur_t	Akin & Karademir 2013	Alfabeya kurdî bi tîpên latînî	11	11
KA_003_kur_t	Akin 2013d	Rêzimana zimanekî bindest	15	15
KA_004_kur_t	Akin & Dilsoz 2013	Nivîskariya ferhengê ya kurdî	14	14
KA_009_kur_m	Akin 2013b	Lêkolîneke înterdisîplîner	13	13
KA_013_kur_m	Akin 2013a	Çend pirsgirekên kurdolojiyê	8	8
KA_019_kur_t	Öpengin 2011	Rewşa kurdî ya sosyolenguîstîk	170	62
KA_020_kur_p	Weqfa Navnetewî 2007	Encamên Psîkolojîk	135	75
KA_021_kur_p	Uzun 1992	Destpêka Edebiyata Kurdî	105	50
KA_022_kur_t	Derince & Mehmed 2012	Perwerdehiya Dînamîk	50	50
KA_024_kur_t	Beşikci & Lezgîn 2008	Ziman – Nasname – Netewe	263	23
KA_025_kur_t	Coşkun et al. 2010	Kula ziman	116	116
KA_026_kur_m	Öpengin 2007	Parastin û guherîna zimên	14	14
KA_030_kur_t	Kurdo 1984	Gramera zmanê kurdi	330	50
KA_045_kur_m	Reşîd 2010	Kurdên Yekîtîya Sovyetê	15	15
KA_046_kur_m	Yüksel 2014	Kurdolojî û Malbata Celîlan	161	46
KA_059_kur_t	Omerxalî & Öpengin 2007a	Destpêk	11	11
KA_060_kur_m	Omerxalî 2007a	Cihwarên Êzdiyên Koçberên	5	5
KA_061_kur_t	Omerxalî & Öpengin 2007c	Sîstema ocax	23	23
KA_062_kur_t	Omerxalî & Öpengin 2007b	Sembolîzma teyran	21	21
KA_063_kur_m	Omerxalî 2007b	Şîrovekirina Sembola Êzdiyan	5	5
Σ			1,490	632

4 Phrasal types of TN-constructions

The concordance consists of 858 tokens: constructions involving a TN. Alongside their frequent employment as subordinators, TNs productively participate in processes of lexical morphology and morphosyntax, thereby resembling any other noun. The data reveal:

- word-internal compounds or derivations ([...]TN[...]), e.g. *weşanên demî* 'periodical publications' (KA_002_kur_t_034), *Serdema Navîn* 'the Middle Ages' (KA_020_kur_p_ 027), *demdirêjî û berdewamiya tedbîrên bikaranîna zimanê zikmanî di perwerdeyê de* 'the longlastingness and continuity of measures for the use of the mother tongue in education' (KA_025_kur_t_122);[12]

- non-determined nouns (TN [...]), e.g. *dem û asta nexweşiyên derûnî piştî bûyerên trawmatîk* 'the duration and level of psychic illnesses after traumatic events' (KA_020_kur_p_083), *berî ku gav biavêjin dibistanê* 'before they set foot in a school' (KA_022_kur_t_034), *mamosteyên wan ji ser mijarê gav kirine* 'their teachers skipped (lit.: stepped over) the topic' (KA_025_kur_t_086);[13]

- determined nouns (DEI TN [...]) involving deictic determiners, e.g. *wê wextê* 'at that time, then' (passim), *gelo* [...] *ev dem çawa tê bibîranîn* 'how [...] this period is remembered' (KA_046_kur_m_038), or TN-INDEF [...], or indefinite determiners, e.g. *weke gaveke duyê jî* 'and secondly/and as a second step' (KA_002_kur_t_019);

- quantified nouns (QUA TN [...]), e.g. *her gav* or *hergav* 'always' (passim), *hin gavên ber bi demokrasiyê ve* 'some steps towards democracy' (KA_020_kur_p_065), *çend wext e* 'it has been some time/for some time' (KA_026_kur_m_052);

- adpositional constructions (ADP [...] TN [...]): *di wê demê de* 'at this period/stage' (passim), *di heman demê de* 'at the same time', *di demên beriya*

[12] As one reviewer correctly observes, these processes are indeed part of lexical morphology, not of (morpho)syntax. As the data show, TNs productively participate in processes at both levels.
[13] This last example is accompanied by a possessive attribute. In the present analysis, this is not counted as a determiner.

Komarê de 'in pre-Republican times' (KA_022_kur_t_040), *ji her wextî zêdetir* 'more than ever' (KA_022_kur_t_035);[14]

- *wh*-phrases (WH [...]), such as *çi demê da* 'in which time(s)' (KA_046_kur_m_055), *çi dema ku* 'whenever/each time' (KA_024_kur_t passim).

Table 2 shows the quantitative distribution of these types in the corpus. More than half of the findings (479 tokens or 56%) are bare nouns. *Dem* as the most frequent type (496 tokens or 58%) has the most varied distribution.

Table 2: Overview of constructions according to phrasal type, absolute figures (total: 858)

Phrasal type	Construction type	*dem*	*gav*	*wext*	*çax*	Σ
bare nouns	TN [...]	220	174	68	17	479
adpositional phrases	ADP [...] TN [...]	191	2	33		226
determined nouns	DEI TN [...], TN-INDEF [...]	58	11	6	9	84
quantified nouns	QUA TN [...]	7	37	2		46
word-internal	[...] TN [...]	14	1	1		16
with *wh*-phrases	WH TN [...]	6		1		7
Σ		496	225	111	26	858

5 Squish 1: Temporal subordination as a subcase of TN modification

With an eye on the manipulation of nominal versus clausal style in academic writing, this section considers various types of constructions in which a TN is in ezafe-construction with a modifying expression. To provide a quantitative context for the qualitative discussion, Figure 1 sets out the proportions of nominally versus clausally modified TNs, with a transitional zone in between.

[14] Adpositional phrases can indeed be determiner phrases on the inside, as an adposition may govern a determined noun. The relationship between the categories can be hierarchical rather than mutually exclusive.

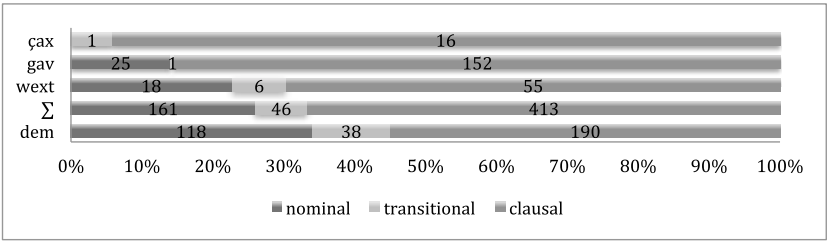

Figure 1: Proportions of nominal, transitional, and clausal TN-modifiers (total: 620)

While the overall tendency points towards predominantly clausal modification of the four TNs (66% or 413 out of 620 modifying tokens being clausal), this tendency is less pronounced in *dem* (54% or 190/346) than in the other expressions: 69% (55/79) for *wext*, 90% (152/178) for *gav*, and 94% (16/17) for *çax*. In what follows, this predominance of clausal constructions will be taken as a reference point from which to observe how patterns transit there from the more nouny areas.

Table 3 below presents the figures for specific constructions, tentatively arranged in terms of a 'categorial squish' (Ross 1972; 2004 [1973]), i.e. along a path of categorial gradience. The squish begins in an area of clearly nominal constructions (25%) built around adjectival, nominal, adpositional and interrogative modifiers of TNs, crossing a transitional area (7%) from action nouns via verbal nouns to participial constructions, and arriving in the internally graded clausal zone (66%).[15] Within the latter, indicative finite clauses without the complementizer *ku* form the endpoint of Squish 1: the *ku*-less usage leaves the full load of the subordinating function to the TN, and the indicative verb form is the farthest possible from any (semi-)nominal status. It is precisely this type of construction that constitutes the most frequent type: 243 or 58% of all clausal findings, 39% of all TN-constructions in the data.[16]

[15] Internal differentiation within the clausy area is considerable, comprising, next to full person marking and the full use of tenses, an indicative/subjunctive distinction, as well as hierarchically complex elaboration on the whole. TN constructions seem to have gone a long way towards sententialisation, projecting structures that expand into areas of rich sententiality. The situation thereby differs from what Lehmann (1988: 193–200) decribes as 'desententialisation', i.e. the loss of – mainly illocutionary – categories as a sentence gets subordinated. Lehmann's concept might more usefully be applied to compare the situation between the clausy and the transitional area.

[16] Constructions of the type 'TN-EZ *ku*...', involving clausal attributes to temporal nouns, can be considered a 'structural bridge' between the transitional zone and the fully subordinating use

Table 3: Constructions based on modified temporal nouns (total: 620)

Level of connectivity	Construction type	*dem*	*gav*	*wext*	*çax*	Σ
nominal	TN-EZ ADJ/NP/poss/AP/wh	118	25	18	–	161
transitional	TN-EZ AN [...]	14		3		17
transitional	TN-EZ VN [...]	22	1	2	1	26
transitional	TN-EZ PAR [...]	2		1		3
clausal	TN-EZ *ku* SBJ clause	16	7	4	–	27
clausal	TN-EZ *ku* indicative clause	78	9	11	1	99
clausal	TN-EZ SBJ clause	9	18	16	1	44
clausal	TN-EZ indicative clause	87	118	24	14	243
Σ		346	178	79	17	620

Nominally modified TN-constructions comprise expressions such as *demên xwe yên dibistanê* 'their own schooldays' (in example (1)), *dema Brêjnev* 'the Brezhnev era' (KA_046_kur_m_025), *wexta Ehmedê Xanî* 'the time of Ehmedê Xanî' (KA_021_kur_p_013f), *dema buhurî* 'past tense' (KA_020_kur_p_010f) etc. Example (2) illustrates how a TN may blend into an adpositional employment, forming an adverbial construction:

(2) ***Dem-a*** *[Sovyet-ê]* *sê* *navend-ên* *kurdnasî-yê*
time-EZ.F Soviet Union-OBL.F three centre-EZ.PL Kurdology-OBL.F
li *Moskva, Lênîngrad* *û* *Yerêvan-ê* *lêkolîn-ên* *dîrok,*
PRP Moscow Leningrade and Erivan-OBL.F research-EZ.PL history
ziman *û* *çand-a* *kurdî* *di-kir-in.* *Ji*
language and culture-EZ.F Kurdish ASP-do.PST-PL CRP
talebext-an *re* *niha tenê çend kurdnas* *wî*
student-OBL.PL CRP now only a.few Kurdologist DEI.OBL.M
kar-î *di-domîn-in.*
work-OBL.M IND-continue.PRS-PL

'In Soviet times, three centres of Kurdology in Moscow, Leningrad and Erivan conducted research on the Kurdish history, language and culture. From among the students now only a few Kurdologists continue this work.' (KA_045_kur_m_007)

of TNs, somewhat resembling a relative clause, as one reviewer suggests. They will be given more attention in connection with the second categorial squish, in Section 6.

Clausal subordination occurs in four basic types, all based on an ezafe-marked form of the temporal noun. Depending on meaning (temporal versus hypothetical, in a rough manner of speaking), the embedded clauses contain an indicative, as in examples (1a) and (3), or a subjunctive verb (4–5). They may or may not be introduced by *ku* (as in examples (3) and (5) versus (1) and (4)); the *ku*-less usage leaves the load of the subordinating function to the TN.[17]

(3) *Li ali-yê din, numûne-yeke din ku*
PRP side-EZ.M other example-INDEF.EZ.F other COMP
di-şibe ne-bûn-a têkili-yê di
IND-resemble.PRS.3SG NEG-exist-EZ.F relationship-OBL.F CRP
navbera mamoste-xwendekar de, ku me berê
space.between-EZ.F teacher-pupil CRP COMP 1PL.OBL before
behs lê kir, di merhale-yeke paştir a
talk PRP.OBL.F do.PST.3SG CRP stage-INDEF.EZ.F later EZ.F
perwerde-yê de tê dîtin ***gav-a*** *[ku*
education-OBL.F CRP IND.come.PRS.3SG see.VN moment-EZ.F COMP
dayik û bav bi tirkî ni-zan-in û ku
mother and father PRP Turkish NEG-know.PRS-PL and COMP
xwendekar jî êdî ni-kar-in derd-ê xwe bi
pupil also any.longer NEG-can.PRS-PL problem-EZ.M REFL PRP
kurdî bêjin].
Kurdish SBJ.say.PRS-PL

'On the other hand, another example that resembles the lack of a relationship between teacher and student, which we mentioned above, is observed at a later stage of the education, when the parents do not know Turkish and the pupils can no longer express themselves in Kurdish.' (KA_025_kur_t_051)

(4) *Ziman ne dijmin, lê destgirê hev in.* ***Wext-a***
language NEG enemy but supporter-EZ.M REC be.PRS.PL time-EZ.F
[feraset-a perwerdehi-ya pirzimanî were qebulkirin],
idea-EZ.F education-EZ.F multilingual SJB.come.PRS.3SG acccept.VN

[17] A fifth type, WH TN-EZ *ku* SBJ clause constructs the temporal noun with a *wh*-element (*çi dema ku... SBJ* 'whenever.../each time...'), expressing contingent iteration. Since this employment is limited to five tokens, all from the same text (KA_024_kur_t), it is presently not considered for the qualitative analysis.

li.ber fikr-a xelet a ku ziman-an wekî
against perception-EZ.F incorrect EZ.F COMP language-OBL.PL as
hevrik (reqîb) û dijmin-ê hev an jî asteng-î li.ber
rival rival and enemy-EZ.M REC or also obstacle-INDEF.EZ against
geşbûn-a ya din di-bîn-e radibe û
development-EZ.F EZ.F other IND-see.PRS-3SG rise.IND.PRS.3SG and
di şûn-ê de nêrîn-a ku di-bêj-e ziman
CRP place-OBL.F CRP vies-EZ.F COMP IND-say.PRS-3SG language
balpişt-ên hev in belav di-ke.
arkadaş-OBL.PL REC be.PRS.PL distribution IND-do.PRS.3SG

'Languages are not rivals, but supporters of each other. When/if the idea of multilingual education is accepted, it rises against the incorrect perception that sees languages as mutual rivals and enemies or as an obstacle to each other's development, and it spreads in its stead the view that says languages are friends and supporters of each other.' (KA_022_kur_t_039)

(5) *Pirani-ya zarok-ên van malbat-an, ji.ber*
majority-EZ.F child-EZ.PL DEI.OBL.PL family-OBL.PL because of
herem-ên koçberi-yê an kurdî-tirkî yanî
region-EZ.PL migration-OBL.F either Kurdish-Turkish that is
duzimanî an jî piranî bi tirkî di-peyiv-in, tenê ji
bilingual or also mostly PRP Turkish IND-speak.PRS-PL only PRP
axftin-ên kurdî fehm di-kin lê bi
conversation-EZ.PL Kurdish understanding IND-do.PRS.PL but PRP
tirkî bersiv-ê di-din. Ji bilî wê weke
Turkish answer-OBL.F IND-give.PRS.PL PRP other DEI.OBL.F as
ku di kom-a din jî de hebû, di vê
COMP CRP group-EZ.F other also CRP exist.PST.3SG CRP DEI.OBL.F
kom-ê de jî ***dem-a*** *[ku di mal-ê de*
group-OBL.F CRP also time-EZ.F COMP CRP house-OBL.F CRP
xizm-ê wek dapîr û bapîr-ên ku bi
relative-EZ.M like grandmother and grandfather-EZ.PL COMP PRP
kurdî di-peyiv-in hebin], ev rewş
Kurdish IND-speak.PRS-PL exist.SBJ.PRS.PL DEI.RCT situation
di-be sedem-a ku zarok ziman-ê xwe
IND-become.PRS.3SG reason-EZ.F COMP child language-EZ.M REFL

yî dayik-ê jibîr ne-kin û
EZ.M mother-OBL.F from.memory neg-DO.PRS.PL and
di-kev-e war-ê teşwîq-ê.
IND-fall.PRS-3SG home-EZ.M encouragement-OBL.F

'The majority of the children from these families, because of the regions of migration, speak either Kurdish-Turkish, i.e. bilingually, or mostly in Turkish; they only understand the Kurdish conversations, but answer in Turkish. Apart from this, as was also the case in the other group, in this group, too, when there is a relative in the house, such as grandmothers or grandfathers who speak Kurdish, this situation becomes the reason why the children do not forget their mother tongue and it becomes a stronghold of encouragement.'
(KA_022_kur_t_020f)

The transitional zone is made up of constructions based on verbal nouns, action nouns and, marginally, participles. Verbal nouns occur in the data in expressions such as *dema xwendina xwe ya îlahiyatê* 'during his theological studies/when he was studying theology' (KA_004_kur_t_066), *dema nefîkirina malbatên kurd* 'during the exilation of Kurdish families/at the time when Kurdish families were exiled' (KA_045_kur_m_003), *dema hatina we-latê mêvandar* 'the time of arrival in the host country' (KA_020_kur_p_018f) etc. Examples (6) and (7), containing verbal nouns, illustrate two internal contrasts: (6) is a junctor-like, (7) a lexical use; the TN-construction in (7) is internally more complex:

(6) ***Çax-ê*** *[axaftın-ê] hewa lı.nav devê*
time-EZ.M speak.VN-OBL.F air between mouth-EZ.M
meriv-an-ra, lı.nav lêv û bêvılê re (k'eop-ra)
person-OBL.PL-CRP between lip and nose-OBL.F CRP #-CRP
dıçe. Bı vi cûre-yi organ-ên (andam-ên)
IND.go.PRS.3SG PRP DEI way-SPC.OBL organ-EZ.PL organ-EZ.PL
axaftın-ê dı-gıhijn-e hev, ya.ji jı hev dûr
speak.VN-OBL.F IND-join.PRS.PL-DIR REC or PRP REC far
dı-k'ev-ın.
IND-fall.PRS-PL

'When speaking, the air passes through the mouth of people, through the lips and the nose. In this way, the speech organs join each other, or they move away from each other.' (KA_030_kur_t_013)

(7) *Têgîn-a çandîbûn-ê weke bi hev re herikîn-a*
concept-EZ.F acculturation-OBL.F as CRP REC CRP flow.VN-EZ.F
rîtuwal û kirin-ên gelerî, tercîh-ên xwarin û
ritual and act-OBL.PL traditional preference-EZ.PL eat.VN and
aktîvite-yan, kompozîsyon-a etnîk ya navkesî ya
activity-OBL.PL composition-EZ.F ethnic EZ.F interpersonal EZ.F
kes-ekî, nirx, nasname-ya tê hîskirin,
person-INDEF.OBL.M value identity-EZ.F come.PRS.3SG feel.VN
guherbar-ên rewş-a koçberi-yê (weke mînak,
variable-EZ.PL situation-EZ.F migration-OBL.F as example
cih-ê ji.dayikbûn-ê, rewş-a nifş di nava
place-EZ.M birth-OBL.F situation-EZ.F generation CRP middle-EZ.F
civak-a mêvandar de, dirêji-ya ***dem-a*** *[mayîn-a*
society-EZ.F host CRP length-EZ.F time-EZ.F stay.VN-EZ.F
welat-ê mêvandar]), û pêwîsti-yan (huner û zext)
country-EZ.M host and need-OBL.PL art and stress
pêk.hati-ye.
be.composed.PAR-BE.PRS.3SG

'The concept of acculturation has been composed as a confluence of rituals and traditional acts, preferences of food and activities, the interpersonal ethnic composition of a person, values, the felt identity, variables of the situation of migration (e.g. place of birth, the generational situation within the host society, the length of the time of stay in the host country), and needs (art and stress).' (KA_020_kur_p_078)

'Action nouns' refer to actions or experiences without being formally derived from verbs, e.g. *êrîş* 'attack', *qir* 'slaughter', *zext* 'stress', *serjimar* 'census', or *perwerdehî* 'education'. Examples (8) and (9), containing action nouns, contrast lexical nouniness (*dema weşana van bernameyan* 'the broadcasting time of these programmes') with nominalised illocution (*mese-leya wextê destpêka perwerdeyê* 'the issue of the time of beginning of the education/the issue of when education should begin').

(8) *19'ê tîrmeh-a 2003'yan, weşan-ên radyo û*
19-EZ.M July-EZ.F 2003-OBL.PL programme-EZ.PL radio and
televîzyon-an ên bi ziman û lehce-yên cuda
television-OBL.PL EZ.PL PRP language and dialect-EZ.PL different
ket-e bin sîwan-eke yasayî. ***Dem-a***
fall.PST.3SG-DIR under umbrella-INDEF.EZ.F legal time-EZ.F

[weşan-a] *van* *bername-yan* *pir* *kurt*
broadcasting-EZ.F DEI.OBL.PL programme-OBL.PL very short
bû *[...]*
be.PST.3SG

'On July 19, 2003, radio and television broadcasting in different languages and dialects came under a legal umbrella. The broadcasting time of these programmes was very short [...].' (KA_025_kur_t_037)

(9) *Li* *ali-yê* *din,* *li.ser* *mesele-ya* ***wext-ê*** *[destpêk-a*
PRP side-EZ.M other PRP issue-EZ.F time-EZ.M beginning-EZ.F
perwerde-yê], *ango* *heke* *li* *dibistan-an* *perwerde-ya*
education-OBL.F that.is whether PRP school-OBL.PL education-EZ.F
kurdî *hebe* *divê* *ji* *çi* *sal-ekê*
Kurdish exist.SBJ.PRS.3SG must PRP which year-INDEF.OBL.F
dest.pê.bike, *hin* *kesan* *got* *ku* *çawa zarok*
begin.SBJ.PRS.3SG some people-OBL.PL say.PST.3SG COMP how child
dest *bi* *dibistanê* *di-kin* *divê* *bi* *perwerde-ya* *bi*
hand PRP school-OBL.F IND-do.PRS.PL must PRP education-EZ.F PRP
ziman-ê *dayik-ê* *dest.pê.bike,* *hinek-an* *jî*
language-EZ.M mother-OBL.F begin.SBJ.PRS.3SG some-OBL.PL also
diyar *kir* *ku* *divê* *zarok hem* *hînî* *kurdî*
statement do.PST.3SG COMP must child both learning-EZ Kurdish
bi-bin *û* *hem jî* *hînî* *tirkî* *bi-bin* *ango*
SBJ-be.PRS.PL and also also learning-EZ Turkish SBJ-be.PRS.PL that.is
divê *perwerde* *bi* *du* *ziman-an* *dest.pê.bike.*
must education PRP two language-OBL.PL begin.SBJ.PRS.3SG

'On the other hand, with respect to the issue of the time of beginning of the education, i.e. concerning those schools that do have Kurdish education, in which year it should begin, some people have said that as children start school, it should begin with education in the mother tongue, and others have stated that the children should learn both Kurdish and Turkish, i.e. that the education should begin in both languages.' (KA_025_kur_t_088)

6 Squish 2: Functional categories of the TN

This section takes a different perspective from the previous one in looking at the occurrence of functional categories within the syntactic domain of TNs: initial ezafe, adpositions, deictic or indefinite determiners, phoric expressions, quantifiers, *wh*, and/or plural marking.[18] These categories, by their nature, can be analysed as indicators of nouniness. Their occurrence can naturally be expected in nominal TN-constructions; what is more interesting about them is how they occur in constructions in which a TN functions as a clause-level junctor. In such syntactic environments, the occurrence of nominal functional categories will be interpreted as a small step back from full junctor status.[19]

As Table 4 shows, only unmodified TNs and nominal TN-constructions are accompanied by the full range of functional categories. However, clausal and transitional TN-constructions can be accompanied by adpositions (52 findings), indefinite determiners (21), as well as, more marginally, *wh* (5), plural (3) and deictic determiners (1).

[18] Possessive attributes/modifiers, as ezafe-constructions, are not counted. They were included in Section 5.

[19] Alberti et al. (2017: 170) speak of "nominal structure build[ing] upon the verbal layers".

Table 4: Nominal functional categories in TN-constructions

Connectivity	Construction type	Initial ezafe	ADP	Deictic det.	Phoric expr.	Quantifier	*Wh*	Indef. det.	Plural	Σ
unmodifed TN		8	95	75	52	51	2	10	1	238
nominal	TN-EZ ADJ/NP/ POSS/AP/WH	2	80	10	–	2	1	54	33	161
transitional	TN-EZ AN [...]	–	7	–	–	–	–	–	–	17
transitional	TN-EZ VN [...]	–	17	–	–	-	-	1	1	26
transitional	TN-EZ PAR [...]	–	2	–	–	–	–	–	–	3
clausal	TN-EZ *ku* SBJ clause	–	–	–	–	–	5	–	1	27
clausal	TN-EZ *ku* indicative clause	–	23	1	-	-	-	20	–	99
clausal	TN-EZ SBJ clause	–	1	–	–	–	–	–	1	44
clausal	TN-EZ indicative clause	–	2	–	–	–	–	–	–	243
Σ		10	227	86	52	53	8	85	37	858

Examples (10–13) illustrate some of these uses. (10) is an adpositional construction based on *dem* and connecting a *ku*-indicative clause; (11), likewise adpositional, works without *ku*. (12) features *dem* with a deictic determiner, connecting a *ku*-subjunctive clause. (13) combines adpositional marking with indefiniteness in forming the expression *di demeke/ê de ku* 'CRP time-INDEF-EZ/OBL CRP COMP', 'while, whereas, at a time when'.[20]

(10) *Dîsa jî* ***di dem-a*** *[ku herêm-a Kurdan ji*
again also CRP time-EZ.F COMP region-EZ.F Kurd-OBL.PL CRP
ali-yê du hukûmet-ên herêmî ve
side-EZ.M two government-EZ.PL regional CRP
di-hat-e rêvebirin], di sal-a 2001'ê ***de***
ASP.come.PST.3SG-DIR govern.VN CRP year-EZ.F 2001.OBL.F CRP
di.nav.de qanûn-ên der.barê 'jiholêkirin-a şerm-ê'
among.these law-EZ.PL concerning remove.VN-EZ.F shame-OBL.F
de hin reform-ên qanûn-a ceza hatin kirin.
CRP some reform-EZ.PL law-EZ.F punishment come.PST.PL do.VN
'Still, in the period when the region of the Kurds was governed by two regional governments, in 2001, some reforms of the penal code, including of the laws concerning the 'removal of shame', were made.' (KA_020_kur_p_050)

(11) *Netîce-yên lêkolîn-eke li.ser encam-ên derûnî û*
result-EZ.PL study-INDEF.EZ.F PRP consequence-EZ.PL psychic and
civakî yên koçberi-ya bi.darê.zorê li.hundir, ku di
societal EZ.PL migration-EZ.F forced internal COMP CRP
sal-a 2002'an da di nav-a jin-ên penaber
year-EZ.F 2002-OBL.PL CRP CRP middle-EZ.F woman-EZ.PL refugee
ên kurd ên li Stenbol-ê de hat-in
EZ.PL Kurdish EZ.PL PRP Istanbul-OBL.F CRP come.PST-PL
meşandin nîşan di-de ku: [...] ji.sedî 90'ê
conduct.VN sign IND-give.PRS.3SG COMP [...] percent 90-OBL.F

[20] While any 'loss of lexicality' (e.g. Lehmann 2015 [1982]: ix) has not so far been an issue in the present study (as the temporal meaning is retained throughout), the comparison between Examples (10), involving *ku*, and (11), without *ku*, nicely illustrates how at the very point of transition between *dem* as a (clausally) modified noun and *dem* in full junctor status ((11) might, of course, still be read as a *ku*-less relative clause), not only the ability to be accompanied by adpositions is retained, but so is the temporal meaning. The phenomena at hand in the present study cannot therefore be regarded as pertaining to issues of 'grammaticalisation'.

wan ***ji*** ***dem-a*** *[koçberî Stenbol-ê*
DEI.OBL.PL PRP time-EZ.F migration Istanbul-OBL.F
bûne] ***û*** ***vir.ve*** *ji pirsgirêk-ên derûnî û*
be.PST.PAR-COP.PL and CRP PRP problem-EZ.PL psychic and
serêş-a timûdaîm gazinc-an di-kin.
headache-EZ.F constant complaint-OBL.PL IND-do.PRS.PL

'The results of a study on the psychic and societal consequences of forced internal migration, which was conducted in 2002 among Kurdish refugee women in Istanbul, shows that [...] 90 percent of them, since (the time) they migrated to Istanbul, have complained of psychic problems and constant headache.' (KA_020_kur_p_041f)

(12) *Li.gorî Salmi, du cure-yên şiddet-a neyekser hene.*
according.to Salmi two kind-EZ.PL violence-EZ.F indirect exist.PL
Yek jê şiddet-a ji.ber îhmal-ê (violence by omission),
one thereof violence-EZ.F PRP omisssion-OBL.F
ev dem-a *[ku mirov di rewş-eke xeternak de*
DEI time-EZ.F COMP person CRP situation-INDEF.EZ.F dangerous CRP
bin yan.jî derfet hebin ku bandor-ên
be.SBJ.PRS.PL or possibility exist.SBJ.PRS.PL COMP influence-EZ.PL
bixisar ên li.ser mirov-an li rû-yê teknîkî
damaging EZ.PL PRP person-OBL.PL PRP face-EZ.F technical
bêne berbendkirin an.jî bêne
SBJ.come.PRS.PL-DIR impede.VN or SBJ.come.PRS.PL-DIR
kontrolkirin, lê ev îmkan neyên bikaranîn],
control.VN but DEI.RCT possibility NEG.come.PRS.PL use.VN
ev şiddet-a ji.ber îhmal-ê ye.
DEI.RCT violence-EZ.F PRP omisssion-OBL.F be.PRS.3SG

'According to Salmi, there are two kinds of indirect violence. One of these, violence by omission, [at] this moment when people are in a dangerous situation or there are possibilities that damaging influences on people can be technically impeded or controlled, but these possibilities are not made use of, this is violence by omission.' (KA_025_kur_t_082)

(13) *Bi vî reng-î ron di-be ku*
PRP DEI.OBL.M colour-OBL.M clear IND-become.PRS.3SG COMP
sebeb-ên bingehî yên betilîn û zelîn-a zimanî eynî
reason-EZ.PL basic EZ.PL attrition and shift-EZ.F linguistic same
ne. Lê.belê, her weku Yağmur (1997:14) îşaret-ê pê
be.PRS.PL but all as Yağmur (1997:14) sign-OBL.F PRP.OBL.F
di-ke, ev diyarde bi wê jî
IND-do.PRS.3SG DEI.RCT phenomenon PRP DEI.OBL.F also
jêk.cuda.dibin ku betilîn-a
be.distinguished.from.each.other.IND.PRS.PL COMP attrition-EZ.F
zimanî diyarde-yeke nava-nifşî ye, ***di***
linguistic phenomenon-INDEF.EZ.F intragenerational be.PRS.3SG CRP
dem-eke de *[ku zelîn-a zimanî pêvajo-yeke*
time-INDEF.EZ.F CRP COMP shift-EZ.F linguistic process.INDEF.EZ.F
di navber-a nifşan de ye].
CRP middle.EZ.F generation-OBL.PL CRP be.PRS.3SG

'In this way, it becomes clear that the basic reasons of language attrition and language shift are the same. However, precisely as Yağmur (1997:14) shows, these phenomena are distinguished from each other by the fact that language attrition is an intragenerational phenomenon, whereas language shift is an intergenerational process.' (KA_019_kur_t_045)

One may assign several of these constructions the status of relative clauses, as one reviewer indeed suggests. However, they do at the same time form part of a categorial continuum that takes TNs from a lexical-nominal to a fully subordinating use. In such a view, *dema (ku)* 'when', *di dema ku* 'at the time when', *ji dema... û vir ve* 'since the time when', *ev dema ku* 'at this time when' and *di demeke de ku* 'at a time when; whereas' can also be analysed as forming a family of morphosyntactically complex and semantically specific (in terms of individuation or deictic discourse anchoring of the temporal situation) subordinators, all based on the TN *dem*. In other words: the functionalisation of *dem* as a temporal subordinator has not yet come to an endpoint of fixed 'grammaticalisation', but can currently be observed at a highly productive stage.[21]

[21] The idea can cross-linguistically be linked to Herkenrath & Karakoç (2017), who, in a typologically different situation, pay attention to the appearance of nominal attributes, determiners, and quantifiers in constructions that otherwise display a range of clausal characteristics; see

7 Intersecting Squish 1 and 2

Squish 1 and 2 intersect where the transitional categories of Squish 1 (action nouns, verbal nouns and, rarely, participles) are constructed with temporal nouns that are accompanied by nominal functional categories. These constructions display traces of nouniness at two levels: in expressing their predicates by means of semi-nominal means (verbal nouns, action nouns, participles) and in featuring some of the said functional categories.

Table 5 positions all analysed examples in their quantitative context with respect to the intersection of Squish 1 and 2. In unmodified TNs, zero, one or two functional categories are equally frequent. At the nominal end of Squish 1 as well as in the more nouny action-noun-constructions within the transitional zone, there is a slight preference for bare TNs without functional categories. However, in the transitional zone of semi-clause-like TN-constructions (based on verbal nouns, and participles), two thirds of the TNs do contain a functional category. Moving into the clausal area, among TN-constructions that contain a finite verb, one finds a clear preference against nominal functional categories. While this tendency is not absolute, the occurrence of functional categories in transitional and clausal constructions seems to be restricted to a single one, and in the clausal area, it is rare overall. The occurrence of twenty clausally modified TNs containing two functional categories, a statistical outlier of sorts, can be exclusively attributed to the fixed expression *di demeke de ku* 'at a time when; whereas', a specialty of two texts and possibly one author/translator. Taken together, this picture can be interpreted as a progressive loss of nouniness as the usage of TNs moves towards clause subordinatorship at the two levels, however, with nouny rebound effects even at advanced stages of clausiness.

also Alberti et al. (2017) on Hungarian.

Table 5: Intersection of Squish 1 and 2 with reference to cited examples

		Squish 2: Number of functional categories				Σ
Squish 1	Construction type	0	1	2	3	Σ
unmodified TN		87 (36%)	61 (25%)	90 (37%)	–	238 (100%)
nominal	TN-EZ ADJ/NP/POSS/AP/WH	77 (47%) Ex. 2	38 (23%) Ex. 1b, 14b	39 (24%)	7 (4%)	161 (100%)
transitional	TN-EZ AN [...]	10 (58%) Ex. 8	7 (41%) Ex. 9, 15	–	–	17 (100%)
transitional	TN-EZ VN [...]	9 (34%) Ex. 6, 7	17 (65%) Ex. 14a	–	–	26 (100%)
transitional	TN-EZ PAR [...]	1 (33%)	2 (66%)	–	–	3 (100%)
clausal	TN-EZ *ku* SBJ clause	26 (96%) Ex. 5	1 (3%) Ex. 12	–	–	27 (100%)
clausal	TN-EZ *ku* indicative clause	75 (75%) Ex. 3	4 (4%) Ex. 10	20 (20%)	–	99 (100%)
clausal	TN-EZ SBJ clause	44 (100%) Ex. 4	–	–	–	44 (100%)
clausal	TN-EZ indicative clause	241 (99%) Ex. 1a	2 (1%) Ex. 11	–	–	243 (100%)
Σ		570 (66%)	132 (15%)	149 (17%)	7 (1%)	858 (100%)

Examples (14) and (15) below are adpositionally governed verbal-noun and action-noun constructions, respectively. (14) coordinates two categorially distinct TN-modifiers: the verbal noun *avabûn* 'founding' (14a) and the lexical noun *kurtejiyan* 'short life' (14b). (15), based on an action noun without any verbal lexical base (*koçberî* 'migration'), has slots for a subject (*jinên kurd ên karker* 'Kurdish working women') as well as for local (*ji Tirkiyeyê* 'from Turkey') and temporal (*di salên 1970'yî de* 'in the nineteen-seventies') information, creating the effect of syntactically complex nominal style that can be seen as a hallmark of contemporary academic writing.

(14) *Komar bi hêvi-yên mezin hat-e vekirin, 1946, lê*
republic PRP hope-EZ.PL big come.PST.3SG-DIR open.VN 1946 but
[...] Lê ***di dem-a*** *[avabûn û kurte.jiyan-a Komar-ê]*
but CRP time-EZ.F found.VN and short.life-EZ.F republic.OBL.F
de, *xebat-eke bêhempa ya çandî çêbû.*
CRP work-INDEF.EZ.F unique EZ.F cultural take place.PST.3SG

'The Republic was opened with big hopes, in 1946, but [...] But at the time of/during the founding and the short life of the Republic, a unique cultural work took place.' (KA_021_kur_p_058)

(15) ***Di dem-a*** *[koçberi-ya jin-ên kurd ên karker*
CRP time-EZ.F migration-EZ.F woman-EZ.PL Kurdish EZ.PL worker
ji Tirkiye-yê di sal-ên 1970'yî] ***de,*** *piranî*
PRP Turkey-OBL.F CRP year-EZ.PL 1970-ADJ CRP majority
jin-ên ciwan ên bi ten-a ser-ê xwe ji.bo
woman-EZ.PL young EZ.PL CRP alone-EZ.F head-EZ.M REFL for
peydakirin-a derfet-ên abor-a xwe bi-kin
find.VN-EZ.F possibility-EZ.PL livelihood-EZ.F REFL SBJ-do.PRS.PL
û ji.bo li xizm-ên xwe yên li welêt
and for PRP relative-EZ.PL REFL EZ.PL PRP home.country-OBL
bi nêr-in koçber bûn.
SBJ-take.care.PRS-PL migrant become.PST.PL

'At the time of migration of Kurdish working women from Turkey in the nineteen-seventies, the majority of young women, who on their own, in order to find possibilities of livelihood and to take care of their relatives back in the home country, became migrants.' (KA_020_kur_p_067)

8 Conclusion

Depending on their syntactic environment, TNs function as nouns, junctors, and adpositions; this allows them to flexibly cross categorial boundaries in shifting between nominal and clausal style, as required in academic writing; their use as subordinators is a special, albeit frequent case. The overall picture reveals a quantitative core within the clausal area, made up of *ku*-less indicative clauses in ezafe-construction with bare *dem* or *gav*. At the other extreme, TNs in unambiguously nouny environments occur on a distinctly lesser scale. The main point of interest of this study have been the quantitative and qualitative squishes and transitions, in an attempt to trace two paths of categorial gradience. Following the first, finite subordinate clauses appear at one end of a scale of TN modifiers, after nouns, action nouns, verbal nouns and participles. Following the second, clause-embedding TNs may express up to two functional categories associated with the NP area. Further, at the intersection of these two continua, constructions can be observed to subtly transit into and out of nouniness at both levels simultaneously.

The four TNs taken together form what can be conceived of as a 'repertoire of items', with the most frequent TN *dem* exhibiting both the most varied use and the lowest proportion of clause-subordinating findings (38% as opposed to 67% in *gav*). While this study has been synchronic[22] in outline and certainly cannot provide a dialectological perspective, differences between authors are discernable, such that, based on a picture of preferences for individual clause-subordinating items, one may arrive at a broad distinction between *dem*-users and *gav*-users, next to some minor other preferential types (see Table 6). These relations are far from conclusive and would require deeper investigation.

[22] With the exception of Kurdo (1991 [1984 [1973]]) and Uzun (1992), all texts are twenty-first-century publications.

Table 6: Frequency of the four TN items by authors[23]

Author(s)/ translator(s)	*dem*	*gav*	*wext*	*çax*	preferred clausal TN
Akin	4/18 (22%)	–	0/1 (0%)	–	*dem*: 4/4 (100%)
Beşikçi & Roşan	14/29 (48%)	–	–	–	*dem*: 14/14 (100%)
Weqfa Navnetewî	28/97 (28%)	1/6 (16%)	–	–	*dem*: 28/29 (96%)
Omerxalî & Öpengin	22/33 (66%)	0/1 (0%)	0/3 (0%)	1/1 (100%)	*dem*: 22/23 (95%)
Akin & Dilsoz	3/19 (15%)	1/1 (100%)	–	–	*dem*: 3/4 (75%)
Öpengin	27/33 (81%)	12/18 (66%)	0/2 (0%)	–	*dem*: 27/39 (69%)
Uzun	0/21 (0%)	17/29 (58%)	0/9 (0%)	4/4 (100%)	*gav*: 17/21 (80%)
Reşîd	3/26 (11%)	10/12 (83%)	–	–	*gav*: 10/13 (76%)
Akin & Karademir	1/13 (7%)	3/4 (75%)	–	–	*gav*: 3/4 (75%)
Coşkun et al.	67/92 (72%)	70/97 (72%)	3/13 (23%)	–	*gav*: 70/140 (50%)
Yüksel & Celîl	13/63 (20%)	17/20 (85%)	17/20 (85%)	0/7 (0%)	*gav/wext*: 17/47 (36%) each
Derince & Mehmet	7/51 (13%)	21/33 (63%)	33/60 (55%)	–	*wext*: 33/61 (54%)
Kurdo	–	0/4 (0%)	2/3 (66%)	11/14 (78%)	*çax*: 11/13 (84%)
Omerxalî	0/1 (0%)	–	–	–	–
Σ	189/496 (38%)	152/225 (67%)	55/111 (49%)	16/26 (61%)	

[23] The item-related columns document the number of clause-subordinating (total: 316) out of the overall TN-usages (total: 858), in absolute figures and percentages, for each item and author (team). The rightmost column lists the subordinating items with the highest frequency among all items employed for this purpose, for each author (team).

References

Aarts, Bas & Denison, David & Keizer, Evelien & Popova, Gergana. 2004. Introduction: The nature of grammatical categories and their representation. In Aarts, B. & Denison, D. & Keizer, E. & Popova, G. (eds.), *Fuzzy grammar: A reader*, 1–28. Oxford: Oxford University Press.

Ágel, Vilmos. 2013. Grammatisches Recycling – im Allgemeinen und im Bereich der Attribute. Vortrag präsentiert auf der Tagung *Attribution, Komplexität, Komplikation*, Justus-Liebig-Universität Gießen, Schloss Rauischholzhausen, 18.–20.09.2013.

Alberti, Gábor & Farkas, Judith & Szabó, Veronika. 2017. HATNÉK-nominalization: Two subtypes of a highly verbal Hungarian deverbal nominalization. *Acta linguistica academica* 64(2). 169–190.

Bedir-Khan, Emir Djeladet & Lescot, Roger. 1970. *Grammaire kurde (dialecte kurmandji)*. Paris: Maisonneuve.

Bolinger, Dwight. 2004 [1961]. Gradience. In Aarts, B. & Denison, D. & Keizer, E. & Popova, G. (eds.), *Fuzzy grammar: A reader*, 311–320. Oxford: Oxford University Press. [Originally published as "The all-or-none", Ch. 1 of: Bolinger, Dwight. 1961. *Generality, gradience, and the all-or-none*. 's-Gravenhaage: Mouton].

Borsley, Robert D. & Kornfilt, Jaklin. 2000. Mixed extended projections. In *The nature and function of syntatic categories* (Syntax and Semantics 32), 101–132. San Diego: Academic Press.

Bresnan, Joan. 1997. Mixed categories as head sharing constructions. In Butt, M. & Holloway King, T. (eds.), *Proceedings of the LFG97 Conferences*, University of California, San Diego. Stanford: CSLI Publications. (`http://cslipublications.stanford.edu/LFG/2/bresnan-lfg97.pdf`) (Accessed 2011-10-07).

Corver, Norbert & van Riemsdijk, Henk. 2001. Semi-lexical categories. In Corver, N. & van Riemsdijk, H. (eds.), *Semi-lexical categories: The function of content words and the content of function words*, 1–19. Berlin: Mouton de Gruyter.

Czicza, Daniel. 2015. Zu Analogien zwischen verbaler und nominaler Organisation. In Hennig, M. & Niemann, R. (eds.), *Junktion in der Attribution: Ein Komplexitätsphänomen aus grammatischer, psycholinguistischer und praxistheoretischer Perspektive* (Linguistik – Impulse & Tendenzen 62), 123–162. Berlin: de Gruyter.

Ehlich, Konrad. 1992. Scientific texts and deictic structures. In Stein, D. (ed.), *Cooperating with written texts*, 201–229. Berlin: de Gruyter.

Fabricius-Hansen, Catherine. 2007. Subjunktor. In Hoffmann, L. (ed.), *Deutsche Wortarten*, 759–790. Berlin: de Gruyter.

Greaves, Chris & Warren, Martin. 2010. What can a corpus tell us about multiword units? In O'Keefe, A. & McCarthy, M. (eds.), *The Routledge handbook of corpus linguistics*, 212–226. London: Routledge.

Heller, Dorothee. 2012. *Wissenschaftskommunikation im Vergleich: Fallstudien zum Sprachenpaar Deutsch-Italienisch* (Arbeiten zur Sprachanalyse 57). Frankfurt am Main: Lang.

Hennig, Mathilde. 2015a. Explizite und elliptische Junktion in der Attribution: Eine Bestandsaufnahme. In Hennig, M. & Niemann, R. (eds.), *Junktion in der Attribution: Ein Komplexitätsphänomen aus grammatischer, psycholinguistischer und praxistheoretischer Perspektive* (Linguistik – Impulse & Tendenzen 62), 21–84. Berlin: de Gruyter.

Hennig, Mathilde. 2015b. Strukturelle Komplexität attributiver Junktion. In Hennig, M. & Niemann, R. (eds.), *Junktion in der Attribution: Ein Komplexitätsphänomen aus grammatischer, psycholinguistischer und praxistheoretischer Perspektive* (Linguistik – Impulse & Tendenzen 62), 163–202. Berlin: de Gruyter.

Hennig, Mathilde. 2016. Einleitung. In Hennig, M. (ed.), *Komplexe Attribution: Eine sprachhistorische, grammatisch-systematische, typologische und funktionalstilistische Annäherung an ein zentrales Nominalstilphänomen* (Linguistik – Impulse & Tendenzen 63), 1–20. Berlin: de Gruyter.

Herkenrath, Annette. 2014. The acquisition of *-DIK* and its communicative range in monolingual versus bilingual constellations. In Demir, N. & Karakoç, B. & Menz, A. (eds.), *Turcology and linguistics* (Éva Ágnes Csató Festschrift), 219–236. Ankara: Hacettepe University Publications.

Herkenrath, Annette. 2015. Junction of attributes in academic Kurmanji: A case of NP-level connectivity. Paper presented at the *6th International Conference on Iranian Linguistics*, 23–26 June, Tbilisi, Georgia.

Herkenrath, Annette. 2016a. A concordance-based look at the semantics of multi-word junctors in Kurmanji. Paper presented at the *3rd International Conference on Kurdish Linguistics*, Universiteit van Amsterdam, 25–26 August 2016.

Herkenrath, Annette. 2016b. Ezafe and the complexity of attribution in academic Kurmanji. In Hennig, M. (ed.), *Komplexe Attribution: Eine sprachhistorische, grammatisch-systematische, typologische und funktionalstilistische Annäherung an ein zentrales Nominalstilphänomen* (Linguistik – Impulse & Tendenzen 63), 295–351. Berlin: de Gruyter.

Herkenrath, Annette. In prep. *KANÎ: Kurmanciya akademîk: Danheviyeke nivîsarên weşandî/Academic Kurmanji: A collection of published texts*. [Pilot corpus in progress].

Herkenrath, Annette & Karakoç, Birsel. 2017. Two questionable candidates for subordinatorship: *-mIşlIK* and *-mAzlIK* in Turkish. *Turkic languages* 21(1). 46–78.

Himmelmann, Nikolaus P. 1992. Grammaticalization and grammar. In *Arbeitspapier 16 (Neue Folge)*. Köln: Institut für Sprachwissenschaft der Universität zu Köln.

Hopper, Paul J. 1987. Emergent grammar. In *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society*, 139–157. Berkeley: University of California.

Hopper, Paul J. & Traugott, Elizabeth Closs. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Kirchner, Mark. 2006. Same source – different paths: Remarks on temporal clauses in Turkish, Azerbaijanian and Persian. In Johanson, L. & Bulut, C. (eds.), *Turkic-Iranian contact areas: Historical and linguistic aspects* (Turcologica 62), 158–164. Wiesbaden: Harrassowitz. [with the editorial assistance of Sevgi Ağcagül & Vanessa Karam].

Kornfilt, Jaklin & Whitman, John. 2012a. Afterword: Nominalizations in syntactic theory. *Lingua* 121(7) [Special issue: *Nominalizations in linguistic theory*]. 1297–1313.

Kornfilt, Jaklin & Whitman, John. 2012b. Introduction: Nominalizations in syntactic theory. *Lingua* 121(7) [Special issue: *Nominalizations in linguistic theory*]. 1160–1163.

Kurdo, Qanatê. 1991 [1984 [1973]]. *Gramera zmanê kurdi: Kurmanci-sorani [Grammar of the Kurdish language: Kurmanji-Sorani]*. İstanbul: Koral. [Kurdish translation from Russian by Kurdo, original title: *Gramatika kurdskogo yzika na matêrialê kurmanci i sorani*].

Lehmann, Christian. 1988. Towards a typology of clause linkage. In Haiman, J. & Thompson, S. A. (eds.), *Clause combining in grammar and discourse*, 181–226. Amsterdam: Benjamins.

Lehmann, Christian. 2015 [1982]. *Thoughts on grammaticalization*. 3rd edn. (Classics in Linguistics 1). Berlin: Language Science Press.

Libert, Alan. 2014. Words which are not conjunctions. Talk given in the University of Newcastle Linguistic Seminar Series, 25 March 2014. (`https://www.academia.edu`) (Accessed 2019-02-14).

Matras, Yaron. 1997a. Clause combining, ergativity, and coreferent deletion in Kurmanji. *Studies in language* 21(3). 613–653.

Matras, Yaron. 1997b. The function and typology of coordinating conjunctions: Evidence from discourse and language contact situations. In Connolly, J. H. & Vimans, R. M. & Butler, C. S. & Gatward, R. A. (eds.), *Discourse and pragmatics in Functional Grammar*, 177–192. Berlin: Mouton de Gruyter.

Matras, Yaron. 2002. Kurmanji complementation: Semantic-typological aspects in an areal perspective. *Language typology and universals (STUF)* 55(1) [Special issue: *Kurdish linguistics*]. 49–63.

Matras, Yaron & Reershemius, Gertrud. 1991. Standardization beyond the state: The cases of Yiddish, Kurdish and Romani. In von Gleich, U. & Ekkehard, W. (eds.), *Standardization of national languages: Symposium on Language Standardization*, 2–3 February 1991 (Unesco Institute of Education Reports 5), 103–123. Hamburg: Unesco-Institut für Pädagogik/Unesco Institute for Education/Institut de l'Unesco pour l'education.

Mithun, Marianne. 1988. The grammaticisation of coordination. In Haiman, J. & Thompson, S. A. (eds.), *Clause combining in grammar and discourse*, 331–359. Amsterdam: Benjamins.

Redder, Angelika. 1990. *Grammatiktheorie und sprachliches Handeln: "denn" und "da"*. Tübingen: Niemeyer.

Redder, Angelika. 2007. Konjunktor. In Hoffmann, L. (ed.), *Deutsche Wortarten*, 483–524. de Gruyter.

Rehbein, Jochen. 1995. Über zusammengesetzte Verweiswörter und ihre Rolle in argumentierender Rede. In Wohlrapp, H. (ed.), *Wege der Argumentationsforschung*, 166–197. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Ross, John Robert. 1972. The category squish: Endstation Hauptwort. In Peranteau, P. M. & Levi, J. N. & Phares, G. C. (eds.), *Proceedings of the Eighth Regional Meeting of the Chicago Linguistic Society*, 316–338. Chicago: Chicago Linguistic Society, University of Chicago.

Ross, John Robert. 2004 [1973]. Nouniness. In Aarts, B. & Denison, D. & Keizer, E. & Popova, G. (eds.), *Fuzzy grammar: A reader*, 351–422. Oxford: Oxford University Press. [First published in 1973 in: Fujimura, Osamu (ed.) *Three dimensions of linguistic theory*. Tokyo: The TEC Corporation, 137–257].

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt Brace.

Szabolcsi, Anna. 1990. Parallelism between noun phrases and clauses. In Bahner, W. & Schildt, J. & Viehweger, D. (eds.), *Proceedings of the 14th International Congress of Linguists*, Berlin, 10–15 August 1987, vol. 3, 2232–2234. Berlin: Akademie-Verlag.

Szabolcsi, Anna. 1994. The noun phrase. In Kiefer, F. & Kiss, K. É. (eds.), *Syntax and semantics*, vol. 27: *The syntactic structure of Hungarian*, 179–274. San Diego: Academic Press.

Thackston, Wheeler M. 2006. *Kurmanji Kurdish: A reference grammar with selected readings*. Cambridge, MA: University of Harvard. (`https://sites.fas.harvard.edu/~iranian/Kurmanji/kurmanji_complete.pdf`) (Accessed 2019-07-08).

Thielmann, Winfried. 2009. *Deutsche und englische Wissenschaftssprache im Vergleich: Hinführen – Verknüpfen – Benennen*. Heidelberg: Synchron.

Appendix

Annotational abbreviations

1	first person	EZ	ezafe	PST	past tense
2	second person	F	feminine	QUA	quantifier
3	third person	FUT	future	RCT	direct case
ADP	adposition	IND	indicative	REC	reciprocal
AN	action noun	INDEF	indefinite	RFL	reflexive
AP	adpositional phrase	M	masculine	SBJ	subjunctive
ASP	aspect	NEG	negation	SG	singular
ATT	attribute	OBL	oblique case	SPC	specificity
CAUS	causative	PAR	participle	SUP	superlative
COMP	complementizer	PHO	phoric expression	TN	temporal noun
COP	copula	PL	plural	VN	verbal noun
CRP	circumposition	POP	postposition	WH	*wh*-expression
DEI	deixis	PRP	preposition		
DET	determiner	PRS	present tense		
DIR	directive	POSS	possessive		

Primary sources/data

Akin, Salih. 2013a. Çend pirsgirekên kurdolojiyê [Some problems of Kurdology]. In Akin, S. (ed.), *Lêkolînên zimannasiya kurdî [Studies in Kurdish linguistics]*, 221–230. Stembol: Avesta. [First published in 2005 in *Kovara Bîr* 1]. [Kurdish original version].

Akin, Salih. 2013b. Lêkolîneke înterdisîplîner li ser bikaranîn û veguhestina navnifşiya kurdî li diyasporayê [An interdisciplinary study on the use and intergenerational transmission of Kurdish in the diaspora]. In Akin, S. (ed.), *Lêkolînên zimannasiya kurdî [Studies in Kurdish linguistics]*, 147–162. Stembol: Avesta. [First presented at the *Konferansa Duyem a Kurdolojiyê*, Duhok, 1–3 May 2011]. [Kurmanji version].

Akin, Salih. 2013c. Pêşgotin [Preface]. In Akin, S. (ed.), *Lêkolînên zimannasiya kurdî [Studies in Kurdish linguistics]*, 11–16. Stembol: Avesta. [Original version in Kurmanji].

Akin, Salih. 2013d. Rêzimana zimanekî bindest çawan tê nivisîn [How to write the grammar of an oppressed language]? In Akin, S. (ed.), *Lêkolînên zimannasiya kurdî [Studies in Kurdish linguistics]*, 39–56. Stembol: Avesta. (First published in 2012 as: Comment écrire la grammire d'une langue minorée? Le cas du kurde kurmanji. In: Arnavielle, T. (éd.), *Voyages grammairiens*, Paris: L'Harmattan, 168–184). [Kurmanji version].

Akin, Salih & Dilsoz, Omer [tr.] 2013. Nivîskariya ferhengê ya kurdî (ferhengnûsiya kurdî) [The writing of a Kurdish dictionary (Kurdish lexicography)]. In Akin, S. (ed.), *Lêkolînên zimannasiya kurdî [Studies in Kurdish linguistics]*, 57–72. Stembol: Avesta. [Kurmanji version, translated from Turkish by Omer Dilsoz].

Akin, Salih & Karademir, Yaqûb [tr.] 2013. Alfabeya kurdî bi tîpên latînî [The Kurdish alphabet in Latin characters]. In Akin, S. (ed.), *Lêkolînên zimannasiya kurdî [Studies in Kurdish linguistics]*, 19–38. Stembol: Avesta. [First published 2006: L'alphabet kurde adapté aux caractères latins. In: Honvault-Ducrocq, R. (direction), *L'orthographe en questions*. Mont-Saint-Aignan: Publ. des Univ. de Rouen et du Havre, 321–333]. [Kurmanji version, translated from French by Yaqûb Karademir].

Beşikci, İsmail & Lezgîn, Roşan [tr.] 2008. *Ziman – nasname – netewe û neteweperwerî [Language, identity, nation and nationalism]*. Stembol: Weşanên Pêrî. [Kurmanji version, translated from Turkish by Roşan Lezgîn].

Coşkun, Vahap & Derince, Mehmet Şerif & Uçarlar, Nesrin & Loran, Elîxan [tr.] & Öpengîn, Ergîn [tr.] 2010. *Kula ziman: Kêseya bikarneanîna zimanê zikmakî di perwerdeyê û serhatiyên xwendekarên kurd li Tirkiyeyê [Scar of tongue: Consequences of the ban on the use of mother tongue in education and experiences of Kurdish students in Turkey]*. Diyarbakır: DİSA. [Kurmanji translation from Turkish by Elîxan Loran & Ergîn Opengîn, original title: Dil Yarası: Türkiye'de eğitimde anadilinin kullanılmaması sorunu ve kürt öğrencilerin deneyimleri].

Derince, Mehmet Şerif & Mehmed, Osman [tr.] 2012. *Perwerdehiya dînamîk a pirzimanî û pirzaraveyî li ser bingehê zimanê dayikê: Modelên perwerdehiyê bo xwendekarên kurd [Dynamic multilingual and multidialectal education based on the mother tongue: Model of education for Kurdish pupils]*. Diyarbakır: Diyarbakır Siyasal ve Sosyal Araştırmalar Enstitüsü/Enstîtuya Diyarbekirê bo Lêkolînên Siyasî û Civakî/Diyarbakır Institute for Political & Social Research (DİSA). [Translated from Turkish by Osman Mehmet, original title: *Anadili temelli çokdilli ve çokdiyalektli dinamik eğitim: Kürt öğrencilerin eğitiminde kullanılabilecek modeller*].

Kurdo, Qanatê. 1991 [1984 [1973]]. *Gramera zmanê kurdi: Kurmanci-sorani [Grammar of the Kurdish language: Kurmanji-Sorani]*. İstanbul: Koral. [Kurdish translation from Russian by Kurdo, original title: *Gramatika kurdskogo yzika na matêrialê kurmanci i sorani*].

Omerxalî, Xanna. 2007a. Cihwarên êzdiyên koçberên cîhanê [Settlements of globally migrating Yezidis]. In Omerxalî, X. (ed.), *Êzdiyatî: Civak sembol rîtûel [Yezidism: Society, symbol and ritual]*, 27–32. Diyarbakır: Avesta.

Omerxalî, Xanna. 2007b. Şîrovekirina sembola êzdiyan: Tawisî Melek [Interpretation of Yezidi symbols: Tausi Melek]. In Omerxalî, X. (ed.), *Êzdiyatî: Civak sembol rîtûel [Yezidism: Society, symbol and ritual]*, 91–96. Diyarbakır: Avesta.

Omerxalî, Xanna & Öpengin, Ergin [tr.] 2007a. Destpêk [Beginning]. In Omerxalî, X. (ed.), *Êzdiyatî: Civak sembol rîtûel [Yezidism: Society, symbol and ritual]*, 33–68. Diyarbakır: Avesta. [Translated from English by Ergin Öpengin].

Omerxalî, Xanna & Öpengin, Ergin [tr.] 2007b. Sembolîzma teyran di êzdiyatiyê de [The symbolism of birds in Yezidism]. In Omerxalî, X. (ed.), *Êzdiyatî: Civak sembol rîtûel [Yezidism: Society, symbol and ritual]*, 69–90. Diyarbakır: Avesta. [Translated from English by Ergin Öpengin].

Omerxalî, Xanna & Öpengin, Ergin [tr.] 2007c. Sîstema ocax, qebîl û esîrên êzdiyan: Termînolojiya wan [The system of ocax, tribe and clan of the Yezidis: Their terminology]. (Ligel lîsteya ocax û qebîlên şêx, pîr û mirîdên êzdiyan). In Omerxalî, X. (ed.), *Êzdiyatî: Civak sembol rîtûel [Yezidism: Society, symbol and ritual]*, 13–24. Diyarbakır: Avesta. [Translated from English by Ergin Öpengin].

Öpengin, Ergin. 2007. Parastin û guherîna zimên: Destpêkek bo çarçoveyên teorîk ên parastin û guherîna zimên û hin împlîkasyonên wan li ser rewşa zimanê kurdî [An approach towards theoretical frameworks of language maintenance and change and some implications for the situation of the Kurdish language]. *Zend* 2007. 52–65.

Öpengin, Ergin. 2011. *Rewşa kurdî ya sosyolenguîstîk li Turkiyeyê [The sociolinguistic situation of Kurdish in Turkey]*. Stembol: Weşanên Avesta. [Translated by Ergin Öpengin from French into Kurmanci, original title: *La situation sociolinguistique du kurde en Turquie: Une étude descriptive et interprétative*].

Reşîd, Tosinê. 2010. Kurdên Yekîtîya Sovyetê: Berê û niha [The Kurds of the Soviet Union: Then and now]. *Bîr* 12. 206–221. (`http://www.kovarabir.com/tosine-resid-kurden-yekitiya-sovyete-bere-u-niha/`) (Accessed 2015-07).

Uzun, Mehmet. 1992. *Destpêka edebiyata kurdî [Introduction to Kurdish literature]*. Ankara: Weşanên Beybûn. [Kurdish version].

Weqfa Navnetewî ya Jinên Azad, Rotterdam and Zanîngeha Utrechtê, Beşa Psîkolojiya Klînîkê û Tenduristiyê and Navenda Agahdariyan a Kurdistanê, Parîs – International Free Women's Foundation, Rotterdam and Utrecht University, Department of Clinical and Health Psychology and Kurdistan Information Office, Paris (ed.). 2007. *Encamên psîkolojîk ên serboriyên trawmayê li ser pêşketina jinên kurd ên penaber li Yekîtiya Ewrûpayê: Encamên dawiyê û dîroka raboriya anketeke li pênc welatên ewrûpî û Tirkiyeyê – Psychological consequences of trauma experiences on the development of migrated Kurdish women in the European Union: Final results and background of a survey in five European countries and Turkey* (Weşanên Weqfa Navneteweyî ya Jinên Azad – International Free Women's Foundation Publications 1). Rotterdam: Weqfa Navneteweyî ya Jinên Azad – International Free Women's Foundation. [Kurdish version].

Yüksel, Metin. 2014. *Kurdolojî û malbata Celîlan: Hevpeyvîn bi Celîlê Celîl û Cemîla Celîl ra [Kurdology and the Celîl family: Interview with Celîlê Celîl and Cemîla Celîl]*. Stembol: Avesta.

9
Kurdish *-râ* as an Anti-Actor marker

Gholamhossein Karimi Doostan & Fatemeh Daneshpazhouh

Abstract: This paper will examine the semantic and syntactic roles of the *-râ* morpheme in Kurdish in the light of Role and Reference Grammar (RRG) (Van Valin, Jr & LaPolla 1997; Van Valin, Jr 2007; Van Valin, Jr & Wilkins 1996). Crucial to RRG is the notion of two semantic macro-roles: Actor and Undergoer. In this work, we argue that the *-râ* morpheme in Kurdish can be analysed as Anti-Actor in the sense that this morpheme is attached to verbal roots in order to prevent their semantic Actor macro-roles from being realized as DPs in the syntax. Moreover, we show that the presence of the *-râ* morpheme in a sentence results in forming a non-active clause in many Kurdish dialects. It is also shown that when *-râ* occurs in intransitive unergative sentences with a compound verb containing an Actor role, it causes the absence of the DPs with Actor roles. Based on this evidence, we claim that the Kurdish *-râ* suffix can only be present when an Actor role is semantically present, but syntactically absent.

1 Introduction

The term *non-active structure* refers to several almost identical structures in which certain external arguments are not manifested as DPs in the syntactic level. Alexiadou & Doron (2012) state that cross-linguistically, there are at least five non-active structures: (1) Anticausatives, i.e. spontaneous events ('break', 'open'); (2) reflexives, which are mostly limited to verbs of body care ('wash', 'comb') and naturally reciprocal events ('meet', 'kiss'); (3) dispositional middles (*This book sells well*), (4) medio-passives; (5) passives. In

Kurdish,[1] a morphologically derived non-active structure is marked by the morpheme *-râ*[2], which attaches to verbal stems. There are two allomorphs of *-râ*: *-rê* and *-râ*, which are sensitive to tense, though the rules are somewhat opaque; cf. Footnote 4 below for discussion. The allomorphs are glossed here with *rê*.PRS and *râ*.PST, respectively. The examples[3] in (1a–4a) and (1b–4b) provide evidence for a distinction between the active and non-active structures in Kurdish respectively.

(1) a. *Sârâ sew-ak-ân=i xwârd.*
Sara apple-DEF-PL=3SG eat.PST
'Sara ate the apples .'

b. *sew-ak-ân xwârd-**râ**-ân.*
apple-DEF-PL eat.PST-râ.PST-PL
'The apples were eaten.'

(2) a. *bâ darka-ka=y dâxist.*
wind door-DEF=3SG close.PST
'The wind closed the door.'

b. *darka-ka ba bâ-i dâx-**râ**.*
door-DEF with wind-OBL close.PRS-râ.PST
'The door was closed by the wind.'

(3) a. *Sârâ samâ=i kird.*
Sara dancing=3SG do.PST
'Sara danced.'

b. *samâ kir-**râ**.*
dancing do.PRS-râ.PST
'Dancing was done.'

(4) a. *Sârâ kitêb-i dâstân bâš da-firoš-e.*
Sara book-EZ fiction well PROG-sell.PRS-3SG
'Sara sells the fiction book well.'

[1] By *Kurdish* in this work, we mean the Sorani Kurdish dialects spoken in the central parts of the Kurdistan regions in Iran and Iraq. The data are mainly from the Mahabadi Kurdish dialect.
[2] This morpheme is represented as *-yâ* in Southern Kurdish dialects.
[3] Abbreviations: 1 = first person, 2 = second person, 3 = third person, OBL = oblique case, DEF = definite, PRS = present, PST = past, EZ = ezafe, PL = plural, PROG = progressive and SG = singular.

b. *kitêb-i dâstan bâš da-firoš-**rê**.*
book-EZ fiction well PROG-sell.PRS-rê.PRS
'The fiction books sell well.'

As it is shown in (1b), attaching *-râ* to the verb causes the omission of the verb's external argument (*Sârâ*); resulting in an intransitive structure. In example (2a), *wind* as a natural force is the subject of the active structure and the clause (2b) shows that attaching *-râ* to the verb causes the elimination of the underlying subject of the clause. This indicates that attaching the *-râ* suffix to the verbal root causes the suppression not only of animate agents but also of inanimate causers. An important issue is the fact that it is possible to attach the *-râ* suffix to intransitive complex predicates consisting of a noun and a light verb as illustrated in (3b). In example (4), the so-called 'middle construction' is formed by adding the *-rê* suffix[4] to the present stem of the verb.

In this paper, following Alexiadou & Doron (2012) and Schäfer (2008), we refer to the sentences in (1b–4b) as non-active constructions. We will examine the syntactic and semantic roles of the *-râ* morpheme in Kurdish in the light of the Actor role in RRG, introduced by Van Valin, Jr & Wilkins (1996), Van Valin, Jr (2007), Beavers (2011) and Beavers & Zubair (2013). This work proceeds as follows: In the next section, we will provide a brief overview of the theory of RRG. In Section 3, the semantic macro-roles, i.e. Actor and Undergoer, of RRG theory will be introduced and slightly redefined. The function of the *-râ* morpheme in various intransitive and transitive constructions will be discussed in Section 4. The final section provides a brief summary.

2 An overview of Role and Reference Grammar (RRG)

In RRG theory (Van Valin, Jr & LaPolla 1997; Van Valin, Jr 2007), the semantic macro-roles have a central and important function, putting the theory in

[4]The general rule for forming the non-active is that it is always based on the present stem, to which *-rê* is added to form the present non-active, and *-râ* is added to form a past non-active. However, there are certain verbs which have an irregularly formed non-active, such as *dâxistin* 'to close', *firoštin* 'to sell', *kawtin* 'to fall', *dân* 'to give', *gutin* 'to say', *girtin* 'to hold', *kirdin* 'to do', *dîtin* 'to see', and especially when the verb base (of the infinitive form) ends with *-din*, such as *xwârdin* 'to eat', as can be seen in (1). Accordingly, not only can *-râ* attach to the present stem (2b, 4b) but it can also attach to the past stem of verbs, as in (1b). For some verbs, it seems that both stems are possible hosts, for example *kir-râ/kird-râ*. This is a topic requiring further research.

a good position to account for the data under investigation. RRG grew out of an attempt to answer two basic questions, which were originally posed during the mid-1970s: (1) What would linguistic theory look like if it were based on an analysis of languages with diverse structures such as Lakhota, Tagalog and Dyirbal, rather than on the analysis of English?, and (2) How can the interaction of syntax, semantic and pragmatics in different grammatical systems best be captured and explained? According to RRG theory, one of the most important ways in which languages differ from each other is the manner in which discourse-pragmatics interacts with the linking between syntax and semantics (Van Valin, Jr 2007: 2). The RRG theory of semantic roles is rather different from that of other theories. According to RRG, semantic roles are assumed to be at work at three distinct levels of generality. At the first level, they are what may be called 'verb-specific' semantic roles, e.g. runner, killer, hearer, broken, etc. At the second level are thematic relations, which are generalizations across the verb-specific roles, e.g. agent, instrument, experiencer, theme, patient. At the third level are generalized semantic roles referred to as semantic macro-roles, i.e. Actor and Undergoer, which are generalizations across thematic relations. *Actor* is a generalization across agent, experiencer, instrument and other roles, while *Undergoer* is a generalization subsuming patient, theme, recipient, and other roles. *Agent* is the prototype for Actor, and *patient* is the prototype for Undergoer (Van Valin, Jr 2007: 53).

The number of macro-roles that a verb takes is generally predictable from its logical structure and there are only three possibilities: 0, 1, 2. If a verb has two or more arguments in its logical structure, e.g. [**do** (*x*, Ø)] CAUSE [BECOME **be-at** (*y, z*)] or **hear** (*x, y*), then the unmarked situation would be to have two macro-roles. If a verb has only a single argument in its logical structure, e.g. **do** (*x*, [**walk** (*x*)]) or BECOME **open** (*y*), then the unmarked situation would be to have only one macro-role. Verbs with no arguments, e.g. (**do** Ø [**snow**]), have no macro-roles. The nature of the macro-roles is also a function of the verb's logical structure. If a verb takes two of them, then they must be Actor and Undergoer. For verbs which have a single macro-role, the default choice follows directly from the logical structure of the verb: if the verb has an activity predicate in its logical structure, the macro-role will be Actor; otherwise, it will be Undergoer (Van Valin, Jr 2007: 62–63).

3 Redefining the Actor and Undergoer

Van Valin, Jr (2007: 61) states that it must be emphasized that the label 'Undergoer' should not be taken literally; the same is true for the label 'Actor'. The Actor of the verb *see* does not do anything, but is nevertheless an Actor in the sense intended here (i.e. the logical subject). The Actor is the participant which is responsible for the state of affairs in the sense that it is impossible to have an action without an entity doing the action, a perceptual situation without a perceiving entity, or a cognitive or emotional situation without a participant experiencing the cognitive or emotional state. Similarly, the Undergoer of a verb like *see* does not necessarily undergo something in the same way as the Undergoer of a verb such as *kill*, but it is still the Undergoer of the verb, i.e. the logical object. In general, the Undergoer represents the non-instigating, affected participant in a state of affairs. Accordingly, Beavers & Zubair (2013: 12) suggests that agentive causation contrasts with non-agentive causation in that in the former, a causer is responsible for causation, while in the latter, a property (state) is the initiator of causation. Van Valin, Jr & Wilkins (1996) assume that the typology of argument properties relevant to 'Agency' assignment in natural languages is as depicted in Figure 1.

We mentioned earlier that Undergoer is a generalized semantic role subsuming thematic roles such as theme and patient, but we posit that the nature of Undergoer is closely related to the notion of affectedness. Affectedness has been linked to many important linguistic domains (Tenny 1994; Beavers 2011). Beavers adopts a two-dimensional space for the encoding of affectedness: the types of change and the degree of change. With respect to the degree of change, Beavers (2011: 2) mentions that, in the following example sentences (5a–5d), the patient (the *apple*) is increasingly less affected.

(5)	a.	John ate the apple up.	(Apple is completely gone)
	b.	John cut the apple.	(Apple cut, not necessarily to a particular degree)
	c.	John kicked the apple.	(Apple impinged, not necessarily affected)
	d.	John touched the apple.	(Apple manipulated, not necessarily impinged)

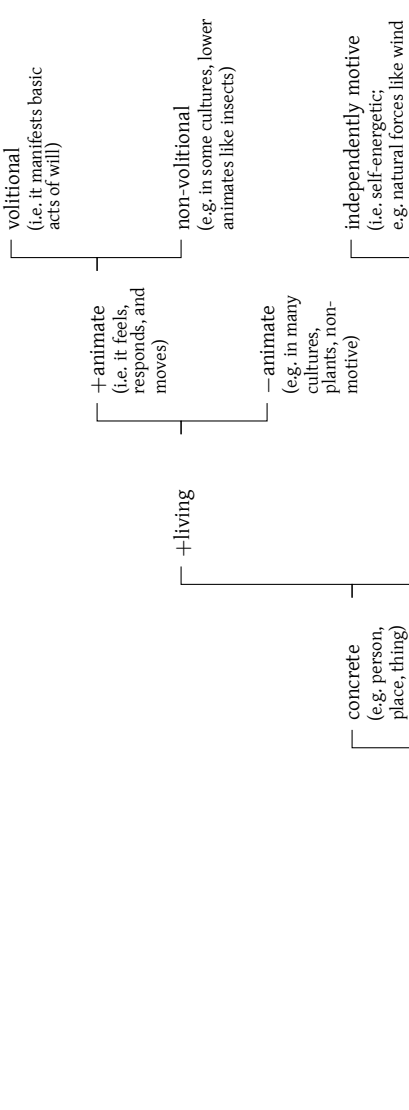

Figure 1: Typology of argument properties relevant to Agency (Van Valin, Jr & Wilkins 1996: 314–315)

Beavers (2011: 5) identifies and introduces the following types of affectedness for some entity *x*, as in (6):

(6) a. *x* changes in some observable property (clean/paint/delouse/fix/break *x*).
b. *x* transforms into something else (turn/carve/change/transform *x* into *y*)
c. *x* moves and stays at some location (move/push/angle/roll *x* into *y*).
d. *x* is physically impinged (hit/kick/punch/rub/slap/wipe/scrub/sweep *x*).
e. *x* goes out of existence (delete/eat/consume/reduce/devour *x*).
f. *x* comes into existence (build/design/construct/create *x*).

Therefore, based on Foley & Van Valin (1984), Van Valin, Jr & Wilkins (1996) and Van Valin, Jr (2007), as well as Beavers (2011) and Beavers & Zubair (2013), we slightly redefine the two semantic macro-roles, i.e. Actor and Undergoer, as follows:

> **Actor:** Actor refers to the first force (*x*) causally involved in the state of affairs. Thus, *x* can be [± entity], [± concrete], [± living], [± animate], [± volitional], [± rational], [± intentional].
>
> **Undergoer:** Undergoer refers to a non-controlling participant (*y*) that undergoes a change (change-of-state, change-of-emotion, creation/consumption,) or undergoes no change (surface contact) in an event.

According to this definition, an Actor role is taken to be all specific roles, e.g. human, animal, natural force, etc., and includes all 'verb-specific' semantic roles, e.g. dancer, runner, killer, etc. Additionally, an Actor role may do something accidentally or intentionally.

4 The semantic and syntactic roles of *-râ*

In this section, we elaborate on the semantic and syntactic roles of the *-râ* morpheme. The function of the *-râ* morpheme in clauses with intransitive complex predicates will be explained in Section 4.1, and the role of the *-râ*

morpheme in clauses with transitive predicates will be elaborated on in Section 4.2. We will extend the analysis of the function of *-râ* to other constructions in Section 4.3. It is worth mentioning that *-râ* cannot be attached to the roots of stative verbs[5] (e.g. 'know', 'hate', 'believe'), so this work is only concerned with eventive or activity verbs. We should also note that in RRG semantic transitivity refers to the number of macro-roles, while syntactic transitivity refers to the number of the direct core arguments. The number of macro-roles is formalized by $[MR_{\alpha}]$, with values $[MR_0]$, $[MR_1]$ and $[MR_2]$.

4.1 Intransitive complex predicates

Sorani complex predicates (CPs) consist of a nonverbal element (a noun, an adjective or a prepositional phrase) and a light verb to form a single predicate. Some of the most common light verbs in Sorani are *kirdin* 'to make/to do', *bun* 'to become', *dân* 'to give', *girtin* 'to hold' and *xwârdin* 'to eat, to collide', as well as *gutin* 'to say'. Intransitive predicates are usually classified into two broad categories based on the thematic roles of their subject arguments: unergative and unaccusative predicates. Unergative predicates (e.g. 'run', 'swim', 'dance') have an agent argument, while unaccusative predicates (e.g. 'fall', 'die', 'bloom') have a non-agent argument. As illustrated in (7–9), the *-râ* morpheme can be attached to the light verbs in unergative complex predicates such as *mala kirdin* ('swim' + 'do') 'to swim', *samâ kirdin* ('dance' + 'do') 'to dance' and *gŏrânī gutin* ('singing' + 'say'). Note that *-râ* cannot occur with simplex unergative verbs (7d).

[5] The reason that *-râ/-rê* cannot attach to stative verbs, as illustrated in (i) and (ii), is because of the fact that these verbs do not have an actor macro-role in their argument structure which could be suppressed by *-râ/-rê*.

(i) a. *Sârâ Kurdi da-zân-e.*
Sara Kurdish PROG-know.PRS-3SG
'Sara knows Kurdish.'

b. * *Kurdi*
Kurdish
*da-zân-**rê**.*
PROG-know.PRS-rê.PRS
'Kurdish is known!'

(ii) a. *Sârâ riq=i la Târâ ya.*
Sara hate=3SG from Tara be.PRS
'Sara hates Tara.'

b. * *la Târâ riq ya-**rê**.*
from Tara hate be.PRS-**rê**.PRS
'Tara is hated!'

(7) a. *Sârâ mala=y kird.* ACTIVE
Sara swim=3SG do.PST
'Sara swam.'

b. *mala kir-**râ*** NON-ACTIVE
swim do.PRS-râ.PST
'Swimming was done.'

c. *Sara pêkan-i.*
Sara laugh.PST-3SG
'Sara laughed.'

d. * *pêkan-**râ**.*
laugh.PST-râ.PST
'Laughing was done.'

(8) a. *Sârâ samâ=y kird.* ACTIVE
Sara dance=3SG do.PST
'Sara danced.'

b. *samâ kir-**râ**.* NON-ACTIVE
dance do.PRS-râ.PST
'Dancing was done.'

(9) a. *Sârâ gŏrânī=y gut.* ACTIVE
Sara sing=3SG say.PST
'Sara sang.'

b. *gŏrânī gut-**râ**.* NON-ACTIVE
song say.PST-râ.PST
'It was sung.'

Comparing the data in (7a–9a) with those in (7b–9b), it is apparent that attaching the non-active marker *-râ* to the verbal roots in (7b–9b) results in the omission of the 'verb-specific' semantic roles such as 'swimmer', 'dancer' and 'singer' or the verb's external argument (*Sârâ*) and thus non-active constructions are formed. This means that the number of macro-roles of each logical structure has been reduced from 1 to 0. Therefore, if a verb has only a single argument in its logical structure, e.g. do '(*x*, [mala'(*x*)]), then by attaching *-râ* to the verb, no macro-role is left for that verb. So, the above non-active structures have no macro-role, but they contain a nonverbal element as an

NP in the subject position. This phenomenon seems to be in line with the Extended Projection Principle (Chomsky 1981) according to which clauses must contain an NP in the subject position.

Interestingly, the *-râ* morpheme can only be attached to unergative complex predicates. This property of the *-râ* morpheme might be related to the nature of transitivity in Kurdish. Haig (2002: 12) states that the notion of transitivity is crucial to the Kurdish verb system. He distinguishes between lexical transitivity, a property of individual simplex lexical verbs, and syntactic transitivity, a property of clauses. He argues that in Northern Kurdish (Kurmanji), CPs such as *can dan* 'die', lit. 'spirit give', e.g. *wî jî can dan* 'He too died', and *dest pē kirin* 'begin', lit. 'hand to.it put' in a clause like *biharē dest pē kirîye* 'Spring has begun' express single-participant events, and entail semantically a single argument. However, the lexical verbs *kirin* 'do' and *dan* 'give' are in themselves lexically transitive, and consequently trigger the ergative construction, regardless of whether an object argument is entailed or not. Thus, according to Haig, unlike syntactic transitivity, lexical transitivity does not necessarily imply a direct object in syntax, but is rather a property of verb roots, reflected in their ability to trigger ergative morphosyntax. Gündoğdu (2016) also proposes that unergative CPs, in Kurmanji Kurdish, are underlyingly transitive structures in which an agentive LV selects a noun element for its nominal object and argues that the noun elements in these CPs are not true direct objects despite fulfilling the argument requirements of the CPs.

It is also worth noting that Sorani displays the phenomenon of tense-sensitive alignment. Alignment in present tenses is nominative/accusative, while the alignment of transitive clauses in the past tense is non-nominative/accusative, and clitics play an important role here. Haig (2008: 289–90) posits that the general rule for clitic placement is that clitics are attached to the leftmost constituents of the verb phrase as in (10), which indicates that, in Sorani, clitics are suffixed to the object (if present) and display person and number features (φ features) of the Actor.

(10) a. *Sârâ sew-ak-ân=i xwârd.*
Sara apple-DEF-PL=3SG eat.PST
'Sara ate the apples.'

b. *Awân sew-ak-ân=yân xwârd.*
they apple-DEF-PL=3PL eat.PST
'They ate the apples.'

The same rule applies in the nominal part of unergative complex verbs. The examples provided in (11) and (12) indicate that the nominal elements, *samâ* 'dancing' and *gorâni* 'singing' in the unergative complex verbs like *samâ kirdin* 'dance' and *gorâni gutin* 'sing' are considered direct object-like arguments in the syntax. In other words, Noun-Verb CPs behave like transitive predicates with respect to the rules of clitic placement, despite the lack of real direct objects in examples like (11–12).

(11) *Ama samâ=**mân** kird.*
we dance=1PL do.PST
'We danced.'

(12) *Awân gorâni=**yân** gut.*
they singing=3PL say.PST
'They sang.'

Therefore, based on the evidence and discussions above, the attachment of the *-râ* morpheme to unergative complex verbs to form non-active voice as in (7b–9b) without the external arguments functioning as subjects in (7a–9a) demonstrate that the presence of the *-râ* morpheme causes the deletion of the external arguments which can be called Actors. We refer to these arguments as Actors because according to Figure 1 they are [± entity], [± concrete], [± living], [± animate], [± volitional], [± rational], [± intentional]. If we are on the right track to call these arguments Actors, then it is plausible to label *-râ* an Anti-Actor morpheme whose presence results in the suppression of Actors. The relationship between *-râ* and the macro-roles in (7–9) is summarized and depicted in the following Table 1.

Table 1: Relationship between *-râ* and Actor macro-role (MR_α) of unergative complex verbs

One-argument predicates	MR_α	thematic relations	semantic roles	verb + *râ*	MR_α
mala kirdin ('swim')	1	do '(*x*, [swim' (*x*)])	*x* = swimmer	*mala kir-râ*	0
samâ kirdin ('dance')	1	do '(*x*, [dance' (*x*)])	*x* = dancer	*samâ kir-râ*	0
gŏrânī gutin ('sing')	1	do '(*x*, [sing' (*x*)])	*x* = singer	*gŏrânī gut-râ*	0

Although, as shown so far, the *-râ* morpheme may appear in unergative CPs, it cannot be suffixed to verbs in unaccusative complex predicates (13–14) such as *naxwoş kawtin* 'get sick', lit. 'sick to fall', and *giyân dân* 'to die', lit. 'soul to give'. However, as example (15) indicates, the light verb of *dân* 'give' can combine with *-râ* if the verb has an activity predicate in its logical structure, where the macro-role is Actor.

(13) a. *Sârâ naxwoš kawt.*
Sara sick fall.PST
'Sara got sick.'

b. * *naxwoš kawt-**râ**.*
sick fall.PST-râ.PST
'It got sick!' (lit.: *Sick was gotten.)

(14) a. *Sârâ giyân=i dâ.*
Sara soul=3SG give.PST
'Sara died.' (lit.: Sara gave up her soul.)

b. * *giyân d(i)-**râ**.*
soul give.PRS-râ.PST
'It was died!' (lit.: Soul was given up.)

(15) a. *Sârâ jinev=i ba min dâ.*
Sara swearing=3SG with me give.PST
'Sara swore at me.' (lit.: Sara gave swearing at me.)

b. *jinev=im pe d(i)-**râ**.*
swear=1SG at give.PRS-râ.PST
'It was sworn at me!' (lit.: Swear was given at me.)

According to our definitions of Actor and Undergoer, the subject arguments in (13a–14a) can be considered Undergoers, but not Actors. The fact that *-râ* in (13b–14b) cannot co-occur with verbs with Undergoer macro-roles (in contrast to what we saw in (7b–9b)) supports our claim that *-râ* is an Anti-Actor morpheme, the presence of which depends on an Actor macro-role. In other words, this Anti-Actor morpheme only appears to suppress an Actor argument as in cases like (7–9); however, it cannot be present when there is no Actor argument to be syntactically suppressed, as in (13b–14b).

4.2 Transitive predicates

As mentioned earlier, in this sub-section we will examine the function of the -*râ* morpheme in clauses containing various transitive verbs with two macro-roles. First, let us consider some clauses with various transitive verbs (16–19):

(16) a. *Sârâ sew-ak-ân=i xwârd.*
Sara apple-DEF-PL=3SG eat.PST
'Sara ate the apples.' CONSUMPTION VERB (ACTIVE)

b. *sew-ak-ân xwârd-**râ**-ân.*
apple-DEF-PL eat.PST-râ.PST-PL
'The apples were eaten.' CONSUMPTION VERB (NON-ACTIVE)

(17) a. *Sârâ nâma-ak=i nūsi.*
Sara letter-DEF=3SG write.PST
'Sara wrote the letter.' CREATION VERB (ACTIVE)

b. *nâma-ka nūs-**râ**.*
letter-DEF write.PRS-râ.PST
'The letter was written.' CREATION VERB (NON-ACTIVE)

(18) a. *Sârâ kteb-ak-ân=ī dit.*
Sara book-DEF-PL=3SG see.PST
'Sara saw the books.' DIRECTED PERCEPTION VERB (ACTIVE)

b. *kteb-ak-ân dit-**râ**-n.*
book-DEF-PL see.PST-râ.PST-PL
'The books were seen.' DIR. PERCEPTION VERB (NON-ACTIVE)

(19) a. *Sârâ tawâw=ī qalam-ak-ân=i dikâr-kird.*
Sara all=3SG pencil-DEF-PL=3SG using-do.PST
'Sara used all pencils.' USE VERB (ACTIVE)

b. *tawâw-i qalam-ak-ân dikâr-kir-**râ**-n.*
all-EZ pencil-DEF-PL using-do.PRS-râ.PST-PL
'All pencils were used.' USE VERB (NON-ACTIVE)

The above examples demonstrate that the verbs such as *xwârdin* 'eat', *nûsin* 'write' and *dikâr kirdin* 'use' involve an action or event with intention and control, whereas *ditin* 'see' does not require any action or effort of any kind on the part of the participant. The active sentences in (16a–19a) indicate that

Sârâ (an Actor) can have all the specific roles of 'consumer', 'creator', 'observer' and 'user' at the 'verb-specific' level. But in the non-active sentences in (16b–19b), the presence of *-râ* collapses verb-specific semantic roles at the syntactic level, suppressing any argument that qualifies as an Actor. This indicates that attaching *-râ* to the two macro-role verbs [MR_2] at the syntactic level causes the reduction of the number of macro-roles by one, changing the [MR_2] to [MR_1]. The relationship between *-râ* and the macro-roles in the transitive verbs under consideration in terms of logical structure argument positions are presented and summarized in Table 2:

Table 2: Relationship between -*râ* and macro-roles of transitive verbs

Active verbs	Samples	MR_α	thematic relation	Semantic roles	verb + *râ*	Semantic roles
consumption	xwârdin ('eat')	2	**do**'(*x*,[**eat**'(*x*,*y*)])	X = CONSUMER Y = CONSUMED	*xwârd-**râ***	*y*=CONSUMED (MR=1)
creation	nûsin ('write')	2	**do**'(*x*,[**write**'(*x*,*y*)])	X = CREATOR Y = CREATION	*nûs-**râ***	*y*=CREATED (MR=1)
directed perception	dîtin ('see')	2	**do**'(*x*,[**see**'(*x*,*y*)])	X = OBSERVER Y = STIMULUS	*dit-**râ***	*y*=STIMULUS (MR=1)
use	dikâr kirdin ('use')	2	**do**'(*x*,[**use**'(*x*,*y*)])	X = USER Y = IMPLEMENT	*dikâr ki-**râ***	*y*=IMPLEMENT (MR=1)

4.3 *-râ* and the Actor role in some other constructions

In the previous sub-sections, the function of the *-râ* morpheme was accounted for in sentences with mainly agentive Actor arguments specified as [+living], [+animate], [+volitional], [+rational], [+intention]. However, we should note that the external arguments, as depicted in (20–21), might be instruments (20a) or natural forces (21a) which, according to Figure 1 and our definition, can be considered Actor Arguments. The examples in (20b–21b) show that suffixing *-râ* to the verbs in (20–21) results in the formation of non-active clauses and the absence of the external arguments, *čaqo* 'knife' and *bâ* 'wind', as direct arguments.

(20) a. *čaqo dast=im=i biri.* ACTOR (INSTRUMENT)
knife hand=1SG=3SG cut.PST
'The knife cut my hand.'

b. *dast=im ba čaqo-e bir-**râ**.* NON-ACTIVE
hand=1SG with knife-OBL cut.PRS-râ.PST
'My hand was cut with a knife.'

(21) a. *bâ darka-ka=i dâxist.* ACTOR (NATURAL FORCE)
wind door-DEF=3SG close.PST
'Wind closed the door.'

b. *darka-ka ba bâ dâx-**râ**.* NON-ACTIVE
door-DEF with wind close.PRS-râ.PST
'The door was closed with wind.'

In addition to the function of *-râ* in suppressing an Actor argument, this morpheme is also able to form middle constructions. A middle construction refers to a clause where the theme or patient of a verb is structurally realized as the subject of the predicate. According to Fagan (1992) and Vendler (1967), the crucial factor for the formation of middles is whether the verb can occur in the present tense, as in *The book reads easily*. Alexiadou & Doron (2012) argue that both passive and middle prevent the insertion of an external argument as subject, but with different properties. They find that cross-linguistically, middle verbs can sometimes be marked as passive (as in Greek) or as active (as in English).

Most scholars agree that the middle construction has the following features that distinguish it from other non-active constructions: (1) middles are generally generic statements, rather than referring to specific events; (2)

middle construction requires the presence of a modifying element, such as an adverbial of manner. The function of *-rê* (the present form of *-râ*) in the formation of middles in Kurdish is illustrated in the following examples:

(22) a. *Sârâ kitêb-i dâstân bâš da-firoš-e.*
Sara book-EZ fiction well PRG-sell.PRS-3SG
'Sara sells fiction books well.'

b. *kitêb-i dâstân bâš da-firoš-**rê**.*
book-EZ fiction well PROG-sell.PRS-rê.PRS
'The fiction books sell well.'

(23) a. *Sârâ tanâf-i bârik da-pičkir-en-ê.*
Sara rope-EZ slender PROG-cut.PRS-CAUS-3SG
'Sara cuts the slender rope.'

b. *Tanâf-i bârik zū da-pičkir-**rê**.*
rope-EZ slender easily PROG-cut.PRS-rê.PRS
'Slender rope cuts easily.'

The sentences in (22b–23b) do not refer to a specific event. They are generic statements containing the manner adverbials of *bâš* 'well' and *zū* 'easy', respectively. What is of significance here is that *-rê* in (22b–23b), like its past form *-râ*, behaves as an Anti-Actor morpheme whose presence causes the deletion of the Actor subject arguments.

Therefore, the data under discussion in Section 4 and the way that the presence of the *-râ/-rê* morpheme changes the argument structure of the clauses suggest the following generalizations:

- If a complex verb has one Actor argument in its logical structure [MR_1], attaching *-râ* to its verbal root causes the absence of its Actor role, resulting in [MR_0].
- If a verb has two arguments in its logical structure [MR_2], attaching *-râ* to its verbal root causes the absence of its Actor role, resulting in [MR_1].

5 Conclusion

This paper has attempted to clarify the semantic and syntactic roles of the *-râ* morpheme in Kurdish in the light of RRG. Crucial to RRG is the notion of

two semantic macro-roles: Actor and Undergoer. Therefore, we first introduced and slightly redefined the Actor and Undergoer macro-roles. Then, we examined the semantic and syntactic functions of the *-râ/-rê* morpheme in various constructions. It has been argued that attaching *-râ* to a verb causes the absence of Actor arguments, resulting in non-active clauses which lack Actor external arguments at the syntax level. Consequently, based on a relatively large and varied body of data, we demonstrated that the presence of *-râ* causes the absence of the arguments with Actor roles (Van Valin, Jr & Wilkins 1996; Van Valin, Jr 2007), in the sentences under study. Accordingly, we suggested that the Kurdish *-râ/-rê* is an Anti-Actor morpheme.

References

Alexiadou, Artemis & Doron, Edit. 2012. The syntactic construction of two non-active voices: Passive and middle. *Journal of linguistics* 48(1). 1–34.

Beavers, John. 2011. On affectedness. *Natural language and linguistic theory* 29(2). 335–370.

Beavers, John & Zubair, Cala. 2013. Anticausatives in Sinhala: Involitivity and causer suppression. *Natural language and linguistic theory* 31(1). 1–46.

Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa lectures*. Dordrecht: Foris.

Fagan, Sarah M. B. 1992. *The syntax and semantics of middle constructions: A study with special reference to German* (Cambridge Studies in Linguistics 60). Cambridge: Cambridge University Press.

Foley, William A. & Van Valin, Robert D., Jr. 1984. *Functional Syntax and Universal Grammar* (Cambridge Studies in Linguistics 38). Cambridge: Cambridge University Press.

Gündoğdu, Songül. 2016. Noun-verb complex predicates in Kurmanji Kurdish: A syntactic account. In Bellamy, K. & Karvovskaya, E. & Kohlberger, M. & Saad, G. (eds.), *Proceedings of ConSOLE XXIII*, 7–9 January 2015, Paris, 383–405. Leiden: Leiden University Centre for Linguistics. (`https://www.universiteitleiden.nl/binaries/content/assets/geesteswetenschappen/lucl/sole/console23-updated.pdf`) (Accessed 2019-08-28).

Haig, Geoffrey. 2002. Noun-plus-verb complex predicates in Kurmanji Kurdish: Argument sharing, argument incorporation, or what? *Sprachtypologie und Universalienforschung (STUF)* 55(1). 15–48.

Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A Construction Grammar approach*. Berlin: Mouton de Gruyter.

Schäfer, Florian. 2008. *The syntax of (anti-)causatives: External arguments in change-of-state contexts* (Linguistics Today 126). Amsterdam/Philadelphia: Benjamins.

Tenny, Carol L. 1994. *Aspectual roles and the syntax-semantics interface*. Dordrecht: Kluwer.

Van Valin, Jr, Robert D. 2007. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.

Van Valin, Jr, Robert D. & LaPolla, Randy J. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.

Van Valin, Jr, Robert D. & Wilkins, David. P. 1996. The case for effector: Case roles, agents and agency revisited. In Shibatani, M. & Thompson, S. A. (eds.), *Grammatical constructions: Their form and meaning*, 289–322. Oxford: Oxford University Press.

Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

10
Revisiting Kurdish dialect geography: Findings from the Manchester Database

Yaron Matras

1 Introduction: Database method and scope

My aim in this paper is to describe preliminary findings from work carried out between 2011–2017 as part of a collaborative project on 'Structural and typological variation in the dialects of Kurdish', based at the University of Manchester. The project's objectives were to create a reference database covering the main areas in which dialects of Kurdish are spoken, to assess typological variation (with particular consideration to possible contact influences), and to investigate the role of verb semantics in the volatility of the ergative construction in Northern Kurdish (Kurmanji and Bahdini). This paper presents findings pertaining to the distribution of structural features, dialect geography, and dialect classification.

The project's data elicitation method was inspired by that used between 2001–2006 to create the Romani Morpho-Syntax (RMS) database (Matras & Elšík 2008; Matras et al. 2009). A questionnaire was prepared in order to capture salient variables in lexicon, phonology and morpho-syntax. Items were translated into second languages that are common in the region (Turkish, Arabic, and Persian). Bilingual speakers were asked to translate the phrases into their local Kurdish dialect. Sessions were recorded and transcribed into templates in which each phrase was pre-tagged for anticipated structures. The data was imported into an open-source database (utilising *MySQL* and *PHP* web interface software), which was made accessible online. It allows the user to filter transcribed phrases by content (Kurdish forms), English elicitation phrase, tags, and speaker's place of origin.

A pilot questionnaire was tested in 2011–2012. It contained around 200 items, of which around half were individual lexemes and function words. The items had been selected based on an assessment of structural variation in samples of connected speech from around 50 recorded interviews of up to 40 minutes each with speakers from various locations in Turkey, Iraq, and Iran, and based on variation documented in existing literature, especially MacKenzie (1961a) but also descriptions of individual Kurdish varieties. Elicitation for the pilot was carried out in a number of locations in the Kurdish speaking regions in southeastern Turkey and northern Iraq and with recent émigrés in Western Europe. The questionnaire was then extended in 2014. The new questionnaire has 300 items and gives special consideration to possible semantic correlates of ergativity, capturing a scale of predicates and participant roles. The approach was inspired by findings on correlates between ergativity, topicality and agentivity in Kurmanji, as presented in Matras (1997) (see also Haig 2008: 215ff.) and in theoretical perspective by Beavers' (2011) semantic analysis of diagnostics for participant affectedness. In addition to the questionnaires, speakers were asked to provide a free speech sample, for which several standardised guideline questions were designed eliciting descriptions of village life, marriage customs, migration, or traditional tales. Free speech samples were generally of 20–40 minutes in duration.

In order to facilitate data collection, project collaborators trained fieldworkers in the region; these were recruited among native speakers who are students of Kurdish language and linguistics at universities in southeastern Turkey and northern Iraq. A protocol was applied by which fieldworkers contacted the project manager based in Manchester with meta-data of proposed speaker consultants and were then given authorisation to carry out recordings, which were archived. The recordings were then forwarded to specially trained native speaker transcription assistants. All questionnaire transcriptions underwent a systematic in-house control and correction procedure by the project team. Sections of 5–7 minutes were selected from each free speech sample for transcription and translation; these transcriptions underwent two consecutive control processes.

Over 200 speaker consultants were recorded, in over 150 locations. The sample shows a bias toward young, educated males, reflecting in part the profile of the fieldworkers and their access to speaker consultants. However, this bias has the advantage of limiting extra-linguistic variability to geographical location. Influence of the Standard language (either Kurmanji or Sorani) has been minimised thanks to the spontaneous elicitation using a second

language as source, but cannot be entirely ruled out; however, the emerging geographical patterns of structural features offer evidence of the non-randomness of speakers' responses. The database, transcribed free speech samples with audio and translation, information on tags/glosses and transliteration symbols, and information on speaker statistics can be found on the project website (`http://kurdish.humanities.manchester.ac.uk`).

On the whole, spontaneous, oral phrase translation has proven to be a reliable method of data elicitation, and convergence to the elicitation (source) language was not found to be an interfering factor. The odd lexical loan from the contemporary contact language (for example, *coz* for 'walnut', from Arabic, in Sabahiya in Syria, rather than the expected *gwîz* as recorded in neighbouring locations) can be taken to represent the free license to incorporate lexical items from the contact language in everyday speech in Kurdish. The same can be said for the occasional repetition, seemingly, of lexical verb forms from the elicitation source phrase, as in *yaşamîş dibim* 'I live' (Turkish *yaşamış*) in several locations in Turkey (among them İmranlı, Pertek, Karlıova, Suruç), *ʃeyş dibim* in Kobane, Syria and *maʃîşe dikim* in Khanaqin, Iraq (both Arabic *ʃîş*), or *zindigî ekem* in Sahneh, Iran (Persian *zendegī*). The fact that the majority of participants – in the case of this particular item, over 90% – opted for translation equivalents that were not direct replications of the item used in the source, but of Iranian-Kurdish etymology – such as *dimînim, dijîm, jiyan dekem*, etc. – suggests that the responses containing a lexical loan reflect actual usage rather than the effect of convergence to the source language. In a small number of cases, some effect of the source language can be detected in the organisation of complex clauses, though the questionnaire is designed to control for such interference by including several sample sentences for each target construction.

Like any targeted elicitation procedure, the method has its limitations. The questionnaire phrases are elicited out of context, and in most cases there were no opportunities to return to the speakers in order to obtain clarifications. Some structures were lost due to mistranslations or other misunderstandings in individual samples, limiting the ability to compare. Funding for the project was obtained thanks to the promise to test a particular formal hypothesis regarding affectedness in transitive constructions (Beavers 2011). This meant that a large portion of the questionnaire had to be devoted to phrases constructed around that hypothesis, at the expense of eliciting other structures (due to the limitation on resources, and therefore the time it would take to record and transcribe data). A number of transcription assistants supported the processing of data, but their work is prone to a va-

riety of influences including standard language norms in both Kurmanji and Sorani, and different levels of experience. Several stages were introduced to control transcriptions for such variation, but inevitably there are some isolated issues that remain. For all these reasons, and others, the database cannot provide a comprehensive overview of all structures that are relevant to the morpho-syntactic typology of Kurdish, and even the comparison among structures that have been elicited will show some gaps.

2 Dialect geography and diffusion centres

Until recently, MacKenzie's (1961a) study of the Kurdish dialects of northern Iraq remained the principal reference work on Kurdish dialect geography. MacKenzie's survey was limited in its geographical scope, covering only around a dozen locations. Nevertheless, the spread of those locations, on either side of the Zabb river, offered a more or less equal level of attention to each of the two dialect groups which MacKenzie named 'Central' (Group 1) and 'Northern' (Group 2). As the most significant phonological difference between the two groups, MacKenzie (1961a: 220–225) notes the shift of Old Iranian inter- and post-vocalic *p* and *m* to *v* in the Northern and *w* in the Central group. The principal morphological isoglosses include the use of enclitic pronominal forms, the presence of a synthetic passive construction, the use of a definite article, the presence of a general plural form, and the loss of case distinction in pronouns in the Central but not in the Northern group; and the presence of nominal case marking, gender and number distinction in nominal attributive endings (Izafe), and a future tense marker in the Northern but not in the Central group. MacKenzie also identifies isoglosses within the Central group. They include the replacement of *ł* by *r* and the retention of grammatical gender in Arbil, Xošnaw, and Rowanduz; the use of the aspectual marker of the progressive-indicative *e-* rather than *de-* and of the so-called 'proximal' demonstrative *em* in Suleimaniya and Warmawa; and some phonetic and phonological specifics. In conclusion, MacKenzie (1961a: 224) proposes a general division between Northern and Central dialects, and a sub-division of the latter between dialects of the Soran-Arbil region to the north, said to be more archaic, and those of the Suleimaniya-Halabja region to the south.

This division of Kurdish into, essentially, three groups – Northern (Kurmanji/Bahdini), Central (Sorani), and Southern (the latter mainly in the Kermanshah and Ilam regions of Iran), with a sub-division of the Central group –

has since been followed broadly in Kurdish linguistics (cf. McCarus 2009). Terminology remains, however, somewhat inconsistent, with the term 'Southern Kurdish' sometimes used as synonymous with 'Sorani' to refer to the 'non-Northern' varieties (see discussion in Haig & Öpengin 2014: 109). However, most current research distinguishes 'Central' in MacKenzie's sense, from 'Southern Kurdish', comprising some of the varieties of the Kermanshah and Ilam regions of Iran (cf. Fattah 2000) and neighbouring regions in Iraq, and parts of the Kordestand and Hamadan provinces in Iran. For the latter, the term 'South Sorani' is also used (cf. Thackston 2006); the precise demarcation of 'Southern Kurdish' remains a matter of ongoing debate (see Belelli, this volume).

MacKenzie's (1961a) discussion of Northern Kurdish (Kurmanji) was limited to varieties of northern Iraq and he was therefore unable to provide any further sub-classification. Öpengin & Haig (2014) address this gap, proposing a geographical sub-division of Kurmanji into five distinct groups. This is based on a selection of features in lexicon, phonology, and verb conjugation. For each group, the authors collected questionnaire data from only one speaker, all originating from Turkey. The classification is flagged as preliminary and the authors emphasise the need for a more fine-tuned investigation, pointing out for example a transition zone southwest of Lake Van. Öpengin & Haig also hypothesise about the classification of Kurmanji varieties in Syria and Iraq, grouping the Bahdini dialects of the Duhok Province of Iraq along with those of neighbouring Hakkari region in Turkey (as Southeastern Kurmanji, SEK), and dividing those of Syria between Southern Kurmanji (SK), which extends to the Hasaka Province of Syria, and Southwestern Kurmanji (SWK), which extends to the Syrian province of Aleppo. Their findings point on the whole to a gradual process of dialect differentiation, especially in lexicon, where the dialects that are farthest apart geographically also share the smallest number of lexical cognates. At the same time, they hypothesise that the totality of Kurmanji and Sorani does not constitute a dialect continuum, with a gradual transition from one extremity to the other. Rather, the division is rather abrupt, with a relatively narrow belt of transitional varieties. The dialects of Hakkari/Duhok (SEK) are a case in point for transitional varieties, showing on the one hand more conservative features than the Kurmanji dialects to the northwest, while on the other hand showing some influences from Sorani.

My approach in this paper is complementary to that adopted by Öpengin & Haig (2014): I draw on data from the Manchester Database survey to reconstruct specifically layers of structural *innovation* and the extent of their diffu-

sion in geographical space, returning then to the question of dialect classification by identifying zones that are the epicentres of such innovations. This approach is based on the assumption that it is innovation that creates differences among related varieties, and that individual innovations differ in the extent of their geographical spread, and so there are no pre-determined dialect boundaries. Rather, the analysis of innovations and their geographical spread can help identify diffusion zones which, put together, can account for the complexity of isogloss intersection (recognising that isoglosses are also subject to stylistic and social variation, as recognised by Öpengin & Haig). This approach draws on the method applied in earlier work on the dialect geography of Romani (Matras 2002; 2005). Consideration is given here to both Sorani (Central Kurdish) and Kurmanji/Bahdini (Northern Kurdish) varieties, including, for the first time, samples of Kurmanji from northern Syria. On the other hand, access to speakers from Iran was limited for both Northern and Southern Kurdish varieties, and therefore few samples were collected for these dialects. All examples can be accessed through the Dialects of Kurdish web resource (Matras et al. 2016); maps are referenced by citing their numbers on the online map index, while transcription examples are referenced 'DB' (Database) and can be consulted online by location, (Kurdish) content and/or English translation.

3 The 'Great Divide' and subsequent innovations

The division between Northern (Kurmanji/Bahdini) and Central (Sorani) reflects two distinct clusters of structural innovations which appear on the map as a dense bundle of isoglosses. Kurmanji/Bahdini innovations include an analytical future tense marker *-ê/-dê/-wê* (Map 3.11.1-3.11.2); an analytical passive construction *tê/hat girtin* 'is/was arrested' (Map 4.2.1, 4.2.2); and reduction of the final clusters **-rd* to *-r* in *kir* 'done' (Map 1.11) and, with the exception of some retention zones, of **-ft* to *-t* in *ket* 'fell' (Map 1.19). Sorani innovations include a definite article *-eke* (Map 3.1.1) and a corresponding plural definite marker *-ekan* (3.1.2); loss of inflection in pronouns, best represented by the absence of a cognate for Kurmanji *ez* 'I' (Map 2.1), and absence of inflected demonstratives (Map 2.4); reduction of gender/number differentiation and (with the exception of some retention zones, see below) emergence of a uniform nominal attributive (Izafe) marker *î* (Map 4.1.1, 4.1.2); reduction of case marking on nouns (Map 3.8.1); a past-tense passive construction *-ra* that can appear either on a light verb or a participle – *desgîr kira*

or *gîra* 'was arrested' (Map 4.2.2); shift from postvocalic *-*v* to -*w* in *aw* 'water' (Map 1.8), *naw* 'name' (Map 1.9) and reduction of final clusters *-*vn* to -*wn/-on* in *kewn/kon* 'old' (Map 1.7) and of *-*ft* to -*wt* in *kewt* 'fell' (Map 1.19).

These innovations (which in some areas tend to cluster on either side of the Zabb river) may be said to constitute a 'Great Divide': They show differences in the internal organisation of paradigms. This can be interpreted as reflecting a prolonged period of tight-knit relations among the respective population groups. This supports the hypothesis of distinct histories of settlement of the two respective groups, as proposed by Jügel (2014), rather than a gradual differentiation *in situ* or even a massive shift in Sorani as a result of admixture with a related substrate, the kind of process suggested by MacKenzie (1961b) in connection with the historical relationship between Sorani and Gorani.

Differences in morphological paradigms and phonology are accompanied by a series of distinct grammaticalisation paths of function words, such as Kurmanji *(li) vê derê/vêrê* vs. Sorani *(l)êre* 'here' (Map 2.5), *niha/anha/nûke* vs. *êsta* 'now' (Map 2.6), *tişt* vs. *hîç* 'anything' (Map 2.10), *hindik/pîçek* vs. *kêmek/tozek* 'a little' (Map 2.11), *pirr/gelek* vs. *zor* 'many' (Map 2.12), and *tişt* (from **tu-şit*) vs. *şit* 'thing' (Map 2.23), as well as distinct lexical items, among them Kurmanji *karim/kanim/şim* vs. Sorani *twanim* 'I can' (Map 2.31), *zarok/biçûk* vs. *mindał* 'child' (Map 2.27), *mezin/fireh* vs. *gewre* 'large' (Map 2.15).

A number of innovations do, however, transcend the Great Divide. In phonology, the retention of *befr* 'snow' (Map 1.23) and the velarisation of *l* > *ł* (Map 1.12) both have their epicentre around Suleimaniya but extend beyond Sorani, the first to the region southwest of Lake Van and up to Muş, the second to the Duhok province in Iraq and beyond to Yüksekova in the Hakkari province of Turkey. The insertion of a 1PL vowel ending in *çûn* > *çûyn* 'we went' (Map 1.2) follows a similar pathway, reaching the Duhok province and the southernmost areas of the provinces of Şirnak and Hakkari. The cluster reduction in *heft* > *hewt* 'seven' (Map 1.18) is still in progress in the Erbil province (around Rowanduz and Khalifan) and reaches the eastern part of the Hakkari province in the north. The spread of *êre* 'here' (Map 2.5) and *gel* 'with' (Map 2.9) has its epicentre similarly in the Suleimaniya area but extends in the north to the provinces of Duhok, Hakkari, and Van. A similar distribution is found for individual lexical items such as *giran* 'expensive' (Map 2.16), while *derga* 'door' (Map 2.25) shows more limited presence in Kurmanji around the easternmost areas of the Duhok province around Akre. The Sorani aspectual ending *-ewe* (e.g. *ew kitêbem xwêndût -ewe* 'I have read this book') appears in the Bahdini dialects of the Duhok province as *-eve* (cf. Sersink *min hevalêd xwe*

dîtn-eve 'I saw my friends', Zakho *jinikê ʕeyne şikandin-eve* 'the woman shattered the mirrors'; DB).

Transcending the Great Divide are also preferences for historically competing lexical options. The retention zone for *pê/pî* 'foot' (Map 2.24) comprises a centre-like area that crosses the Divide, contrasting with diverse lexical innovations in the peripheries. A somewhat comparable picture, though narrower in geographical spread, is the emergence of related forms for the 2PL pronoun *hing* in the eastern part of the Duhok province (Akre) and in Şemdinli in the neighbouring Hakkari province, and *engo* in the northern part of Erbil province (Rowanduz, but extending to Khalakan) and south of Lake Urmia (Mahabad, Oshnaviyeh), whereas the peripheries have *hûn/wen* (Kurmanji) and *êwe* (Sorani) (Map 2.2).

There are also further cases of convergence between Northern and Central Kurdish: Sorani generally has *sewz* 'green' but the Kurmanji form *kesk* extends to Erbil province (Map 2.14), and the form *kengê* 'when', which is common in Kurmanji is also found as far south as Khalakan in Iraq and Sardasht in Iran, contrasting with *kê* which is more predominant in Sorani (Map 2.7). All this supports Öpengin & Haig's (2014) observation that the Kurmanji frontier dialects are subject to Sorani influence, but also the possibility of a two-way convergence area, as proposed by Jügel (2014).

4 Epicentres and diffusion of innovations

Within each side of the Great Divide we can identify additional innovations that do not extend to the group in its entirety but are distinctive in their distribution of particular sub-areas. A **Western Kurmanji** innovation zone encompasses the area west of Muş, from Gaziantep in the south to Erzurum in the north. A defining feature of this area is the spread of the adjectival demonstrative form *va* (Map 2.3), the future tense in *ê* (Map 3.11.3), a strong tendency toward simplification of the nominal attributive (Izafe) plural marker to *-ê* (Map 4.1.2), a tendency toward loss of the pharyngeal in *haywan* 'animal' (Map 1.27), lexical preferences like *ning* 'foot' (Map 2.24), and incipient tendencies toward diphthongisation in *heyşt* 'eight' (Map 1.1), reduction of the final cluster in *kevn* > *kewn* 'old' (1.7), and the analytical formation *çi wextê/çi çax* 'when' (Map 2.7). Several developments are contained in the westernmost area of this zone and might be considered to be more recent: the stem consonant in *kanim* 'I can' (Map 2.31), the analytical formation *deh û pênc* 'fifteen' (Map 2.18), and *çitan* 'how' (Map 2.8). By contrast,

on the fringes we find several clusters of regionally contained innovations in areas that are otherwise by and large coherent with Western Kurmanji: A central area (between Diyarbakır and Varto) shows loss of oblique case marking in *bajêr* > *bajar* (Map 3.5.3), preference for the double oblique construction with past-tense transitive predicates (Maps 4.7.1– 4.10.2), the form *anha* 'now' (Map 2.6), and acquisition of pharyngealisation in *ḥeşt* 'eight' (Map 1.28). A southernmost area around Qamishli/Nusaybin/Kızıltepe shows a future tense marker *wê* (Map 3.11.3), prevalence of *c(iy)a min* 'my mother' (Map 2.21), *çilo* 'how' (Map 2.8), *piçêk* 'a little' (2.11), *duduwa* 'second' (Map 2.19), reduction of the postposed marker *ra* > *r* (DB), and directional preposition *cem* (Map 3.6.1). Finally, an area to the northeast (between Tatvan, Eleşkirt, and Doğubeyazıt) shows insertion of a glide in *gweh* 'ear' (Map 1.32) and use of *çankî* 'how', shared with the area around Lake Van to the south (Map 2.8).

At the other end of the Kurmanji dialect continuum, we can identify a dynamic **Southeastern Kurmanji** innovation zone with its epicentre in the Duhok province extending northwards to Hakkari province, reaching Yüksekova in the east, to the provinces of Muş and Van in the north, and to Hasaka in the east. Distinctive features include the fronting of the vowel *û* to *î*, a process that is hierarchical in its progression, with *hemû* > *hemî* 'all' (Map 1.6) showing the widest distribution, reaching the provinces of Hakkari, Van and Muş (Turkey) as well as Hasaka, followed by *bû* >*bî* 'was' (Map 1.4), with a similar reach but greater variability, while *dûr* > *dîr* (Map 1.3) is more regionally contained, with wider distribution of an intermediate form *dür*. Further developments include metathesis in *berf* > *befr* 'snow' (Map 1.23) and the analogous replication of a final stop in *bab-* 'father' (Map 1.10), future tense marker *dê* (Map 3.11.3), prevalence of *dayka min* 'my mother' (Map 2.21) and *(di)gel* 'with' (Map. 2.9) as well as *piçêk* 'a little' (Map 2.11) and *biçûk* 'child' (Map 2.27). More contained, extending to the neighbouring Hakkari province but not to Van, is the velarisation of *ł* (Map 1.12) and use of *şim* 'I can' (Map 2.31), while limited to just the Duhok region are the plural nominal attributive marker (Izafe) *-êd* (Map 4.1.2), absence of an overt relative clause marker (Map 4.3.1), and use of *nûka* 'now' (Map 2.6), *çi* 'anything' (Map 2.10), and *duwê* 'second' (Map 2.19), which latter extends eastwards to Hasaka province in Syria. The northernmost area also shows some features that are not shared with the Duhok province, such as the syllable structure in *(ʕ)ezman* 'language' (Map 1.20).

On the Sorani side of the Great Divide, we can similarly identify two principal innovation zones, as noted by MacKenzie (1961a). The Suleimaniya province is the epicentre of a **Southern Sorani** innovation zone that fea-

tures the shifts **kewn* > *kon, kun* 'old' (Map 1.7) and **mizgeft* > *mizgewt* 'mosque' (Map 1.17), which extend to Lake Urmia in the north and partly to the Erbil province (Rowanduz, and farther north to Khalifan), and generalisation of enclitic pronouns as possessive markers of the type *mał-im/mał-eke-m* 'my house' (Map 4.1.3-4.1.5), extending to the southern part of the Erbil province but only sporadically north of Khalifan, where the analytical type *mał-î min* (often gender-inflected) prevails. The demonstrative *em* (Map 2.3), 2PL pronoun *êwe* (Map 2.2), the forms *çon* 'how' (Map 2.8), *tozek* 'a little' (Map 2.11), *tir* 'other' (Map 2.13), the preposition *bo lay* 'to' (Map 3.6.1) and reduction of the 1SG pronoun *emin* > *min* (Map 2.1) also have their epicentre in the Suleimaniya zone, extending to Lake Urmia but not (or only sporadically) to the Erbil province. Forms like *kêy* 'when' (Map 2.7) on the other hand are shared primarily with the southern part of the Erbil province. More contained within the zone are the reduction of the cluster *nd* to *n* in *minał* 'child', *dewlemen* 'rich' (Map 1.2.1, 1.2.2), preference for indicative progressive in *e-* (Map 3.10.1, 3.10.2), and distinctive lexical items like *qaç* 'foot' (Map 2.24).

A **Northern Sorani** innovation zone extends from the area between Erbil, Rowanduz, Khalakan, and Mawat in Iraq, and across to Mahabad, Oshnaviyeh, and Urmia in Iran. Many of its shared developments seem to be incipient and subject to considerable variability: Processes of palatalisation affecting different word positions, as in *guh* > *cuh* 'ear' (Map 1.32), *kenge* > *kence* 'when' (Map 1.35), *nezikî* > *neziçî* (Map 1.34, cf. Map 1.35), incipient depalatalisation in *kiç* > *kits* (Map 1.16), and pharyngeal substitution *ḥ* > ʕ and ʕ > *ḥ* (Map 1.24, 1.25, 1.26), as in Erbil *ʕefte u ʕevd* 'seventy seven', Shaqlawa *ʕazir* 'ready', Choman *ʕapis* 'prison', Khalifan *ḥeşîret* 'clan, tribe', Piranshahr *ḥereb* 'Arab'. The analogous replication of a final stop in *bab-* 'father' (Map 1.10) is found here too, linked with the Southeastern Kurmanji area across the Great Divide. Distinctive of the zone is the 2PL pronoun *engo* (Map 2.2), similarly related to its counterpart *hing* immediately across the Great Divide, as well as the form *dîke* 'other' (Map 2.13). Contained within the area of northern Erbil province is the substitution of liquid consonants *mał* > *mar* (in some cases possibly from a proto-form **lr*) (Map 1.12) and the form *kû* 'how' (Map 2.8). The varieties on the Iranian side of this innovation zone are known as Mukri (Öpengin 2016). Arguably, their distinctive character is a product of sharing some innovations with Northern Sorani that do not extend south to the Suleimaniya province, and others with Southern Sorani that do not extend to the northern sections of the Erbil province around Rowanduz and Khalifan. Like other sectors of Northern Sorani, Mukri too is also a retention zone (see below), which again makes it distinct from the varieties to the south. Distinc-

tive lexical items include *çêw* 'mountain' (Map 2.28) and *laq* 'foot' (Map 2.24), also shared with some varieties to the south, while a unique innovation is the emergence of an analytical progressive aspect: Mahabad *le hałi xwêndini kitabe*, Marivan *xerîkî xwendinewey kitabe* 'he is reading a book' (DB).

5 Retention zones

The absence of shared innovation is, in historical perspective, a weak indicator of the cohesion of a regional speech community and therefore of lesser diagnostic value for dialect groups (cf. Matras 2002: Ch. 9), yet the dialect landscape does feature a number of retention zones, which contribute to the distinctive character of some regional varieties and of course help define isoglosses between them. Retention of nominal case marking follows a hierarchy: Kurmanji varieties generally retain the oblique case on feminine nouns but only in some masculine nouns (e.g. nom. *bajar* 'town', obl. *bajêr*, but note the retreat in some areas – see above). The Southeastern Kurmanji zone is also a retention zone for the oblique case marker *-î* on masculine nouns, as in the directional object obl. *gund-î* 'village' (Map 3.4.1, 3.8.2) and the past-tense transitive subject *ẓełam-î ker dikêṣa* 'the man was pulling the donkey' (Map 3.2.1). A core area within the adjoining Northern Sorani zone shows a tendency toward retention of an oblique suffix *-î/-y* in determined objects, as in *emin ew piyawe-y/jine-y denasim* 'I know this man/woman' (Map 3.8.1, 3.8.2), absent elsewhere in Sorani. The loss of gender distinction in nominal attributive endings (Izafe) is widespread in Sorani but is retained for some nouns in some of the same sectors within Northern Sorani, e.g. Khalifan *bawç-ê min* 'my father', *dayk-a min* 'my mother'; *xaniy-ê min* vs. *mar-a min* 'my house' (DB). Analytical possessive pronouns are similarly shared between Kurmanji as a whole and the Northern Sorani conservative retention zone, cf. Qalat Diza and Sardasht *daykî min* 'my mother' (Map 4.1.3, 4.1.4, 4.1.5).

In the verbal system, the historical 3SG ending *-t* survives in selected verbs, most notably 'to come', in a retention zone covering Southeastern Kurmanji and Northern Sorani – for example, Şemdinli (Hakkari province) and Sersink (Duhok province) *tê-t* 'he is coming', Rowanduz, Erbil, and Marivan *dê-t* – and is optional in some of the other Sorani dialects as well, cf. Suleimaniya *yê-t* (DB). Our maps document this form in examples such as Zakho *jinkê ḥeskir lawja bêj-ît* 'the woman wanted to sing', *kurikê biçîk kitêbê naxwîn-it* 'the small boy is not reading the book' (Map 3.9.2, 3.9.3). Retention of canonical ergativity (nominative marking of the direct object and verb agreement with the

object in past-tense transitive clauses) is a conservative feature within Kurmanji and the construction remains least eroded in the Kurmanji peripheries, especially in the southeast (Duhok province). Sorani as a whole would constitute a retention zone with respect to the synthetic passive, if the form in *-r-* is a retention of the Indo-Iranian predecessor, as in Sangaw *ekuj-r-ên* 'they are killed' (Map 4.2.1).

6 The Kurmanji dialects of Syria

Documentation of the Kurmanji dialects of Syria has been lacking until recently. Speakers conventionally divide these dialects into three groups: According to Ahmed (2016), Aşîtî varieties are spoken between the Iraqi border and the eastern suburbs of the city of Qamishli; Ẋerbî is spoken between Qamishli and the border between Hasaka and Raqqa provinces to the west; and Afrînî is spoken in Syria between Raqqa province, Kobane and Afrin, to the west, though speakers often regard the varieties of Kobane and Afrin as distinct dialects. Ahmed (2016) suggests that the three dialects of Syria may be related to the three-way division of the Kurmanji dialects of Turkey proposed by Haig & Öpengin (2018). Table 1 presents a selection of items from the Manchester Database that document four locations from northern Syria, arranged from west (left) to east (right), and compares them to data from neighbouring Zakho in Iraq.

As Table 1 clearly shows, the four Syrian Kurmanji varieties form a dialect continuum, not just among themselves but also in relation to the variety of neighbouring Zakho in Iraq. The Table nicely illustrates the hierarchical spread of the fronting of *û* to *î* from east to west, with Zakho showing *hîn* 'youPL', *bî* 'was' and *hemî* 'all', Derik showing fronting only in *bî* 'was' and *hemî*, and Qamishli only in *hemi*. Features shared between Zakho and Derik include the absence of diphthongisation in *ḥeşt* 'eight' and the forms *duwê* 'second' and *giran* 'expensive', while otherwise a cluster of isoglosses separates the Zakho dialect from those of Syria. From the selection of items in the Table no particularly close affinity stands out between the dialects of Afrin and Kobane, both known as *Afrînî*, and this represents the general picture for the two samples in the Manchester Database. The Kobane variety in fact shares a series of features with dialects recorded in Turkey (both in the Manchester Database and as reported by Öpengin & Haig 2014 and Haig & Öpengin 2018, among them lexical items such as *qîzik* 'girl', a preference for light verb construction with loans as in *ʕeyş dibim* 'I live', initial glottal

Table 1: Comparison of selected forms for four Syrian Kurmanji varieties, and Zakho in Iraq

	Basselhâya (Afrin)	Kobane	Qamishli	Derik	Zakho
you.PL	*hûn*	*hûn*	*ûn, win*	*hûn*	*hîn*
was	*bû*	*bû*	*bû*	*bî*	*bî*
all	*ḥemû*	*ḥemû*	*hemi*	*ḥemî*	*hemî*
far	*dûr*	*dûr*	*dûr*	*dür*	*dür*
eight	*ḥeyşt*	*ḥeyşt*	*ḥeyşt, ḥeşt*	*heşt*	*ḥeşt*
second	*dudu*	*duduya*	*duduwa*	*duwê*	*duwê*
expensive	*biha*	*biha*	*biha*	*giran*	*giran*
other	*din*	*din*	*di*	*dî*	*dî*
girl	*keçik*	*qîzik*	*keçik*	*keçik*	*keçik*
how	*çawa*	*çawa*	*çawa, çilo*	*çawa*	*çawa*
small	*piçûk*	*çûçik*	*piçûk*	*kiçik*	*biçîk*
walnut	*gûz*	*gwîz*	*gwîz*	*guze*	*gîz*
these things	*evan tişta*	*va tiştana*	*ev tişt*	*ev tiştê ha*	*ev tişte*
noon	*nîvro*	*nîvro*	*nîro*	*nîro*	*nîvro*
today	*îro*	*hîro*	*îro*	*îro*	*ev roke*
I eat	*dixwim*	*dixum*	*dixwim, dixum*	*dixum*	*dixwim*
my mother	*dayka min*	*diya min*	*ca mi*	*ciya min*	*deyka min*
foot	*ling*	*nig*	*nig*	*ning*	*pê*
now	*aniha*	*niha*	*ana*	*neha*	*nûke*
I can	*kanim, karim*	*kanim*	*karim*	*karim*	*dişim*
Izafe pl.	*-ê*	*-ên*	*-ê*	*-ê*	*-êd*
to the town	*bajêr*	*gund*	*bajar, bajêr*	*bajar*	*gundî*
I live	*diʃeyişim*	*ʃeyş dibim*	*diʃeyişim*	*diʃeyişim*	*dijîm*
I work	*îş dikim*	*îş dikim*	*dişuẋulim*	*dişuẋulim*	*kar dikim*

in *hîro* 'today', and the demonstrative *va*. As noted above, the area around Qamishli and neighbouring Nusaybin and Kızıltepe in Turkey shows a number of distinctive innovations. It follows that Syrian Kurmanji fits in nicely within the West-to-East continuum of Kurmanji dialects, its feature distribution reflecting both the somewhat interrupted settlement pattern of Kurds in northern Syria and their strong ties with communities on the other side of the Turkish and Iraqi borders rather than a separate status as a coherent dialect periphery.

7 Conclusion

The findings outlined above, based on the largest-scale survey to date of Kurdish dialects, confirm, broadly speaking, observations made by MacKenzie (1961a), Öpengin & Haig (2014) and Haig & Öpengin (2018) on the basis of much smaller samples: There is a well-pronounced divide between Kurmanji and Sorani, and sub-groups dividing Sorani into a Northern and Southern sector, and Kurmanji into a Western, a South(east)ern, and a transitional zone (note that Haig & Öpengin (2018) revise their earlier classification into five groups, merging them into three). The method proposed here, however, differs from those employed in the other studies, and this has some implications. First, rather than adopt a deductive approach by postulating dialect groups on the basis of pre-selected locations or speakers, thus running the risk of a pre-defined nomenclature of dialect classification, the method adopted here is inductive, as it searches for patterns within a wide-coverage survey and then identifies a classification based on the attested connections among clusters of samples and data points. Second, by distinguishing innovations from retention zones, and acknowledging the hierarchical nature of innovations in regard to 'depth', stability, and territorial spread, we obtain a dynamic understanding of historical differentiation rather than rely on a static snapshot of dialect differences.

The picture that emerges is that of a) four principal diffusion centres or innovation zones, b) two principal and adjoining retention zones on either side of the Great Divide, and c) a tendency for some Southern Sorani innovations to reach the southernmost Kurmanji varieties. Haig & Öpengin (2014: 108) propose that Kurdish (as a whole) is not a typical dialect continuum that results from the gradual spread from a common geographic source but the outcome of two initially distinct groups speaking closely related varieties, with subsequent contact among them. Jügel (2014), in effect, puts forward

the same view. Both studies attribute a possible role to language contact: The former speculates about an Armenian substrate in Kurmanji, the latter about a Semitic sub- or adstrate in Sorani. The movement of populations speaking related varieties who have migrated and settled across the region is of course well attested, if we consider the dispersal of speakers of Zazaki, Gorani, Feyli and *Şêx Bizînî* Sorani in central Anatolia, as well as of speakers of other, non-related languages, such as Domari, Neo-Aramaic, or Azeri (Turkmen). This makes the hypothesis of two distinct groups settling in proximity to one another plausible.

The focus on innovation zones and their diffusion centres leads us to hypothesise the following historical scenario of dialect differentiation in Kurdish: Stage 1 sees the settlement of two groups with related but distinct speech varieties on either side of the Zabb river. The two varieties differ primarily in alignment structures (Proto-Kurmanji relying on inflection while Proto-Sorani relies on clitics) and in the interplay of nominal case and definiteness (Proto-Kurmanji being case-oriented while Proto-Sorani is deixis-oriented). The two varieties also differ in some (albeit few) phonological features and in lexical features that arise either from distinct selections among historical options or, in the case of function words, from different grammaticalisation paths. In Stage 2, following settlement and possibly geographical expansion, two dynamic innovation centres emerge on each of the peripheries – Western Kurmanji and Southern Sorani. As Southern Sorani drifts further away from its ancestor variety, losing all gender and case marking and relying on pronominal clitics as possessives, Northern Sorani continues to retain some conservative features in nominal morphology that bear similarities to the adjacent Kurmanji dialects. By Stage 3, the two populations, possibly as a result of further expansion, intensify contacts in the area around the Zabb river. As a result, some innovations originating from the Southern Sorani diffusion zone, including some that fail to spread into Northern Sorani, reach the Southeastern Kurmanji varieties. These varieties, in turn, susceptible to contacts from the south, develop into an innovation zone in their own right and exert influence on neighbouring Kurmanji varieties to the north, extending up to Lake Van, a phase that we might categorise as Stage 4. Finally, at Stage 5, innovations emerge that are still incipient and more regionally contained, shaping the 'central' Kurmanji transition zone and peripheries to the north and south, and further strengthening the cohesion of Northern Sorani. The outcome is the present-day complexity of intersecting isoglosses that reflect larger-scale spread of innovations, conservative retention zones, and more localised developments.

Acknowledgements

The present paper draws on research enabled by a small grant from the British Academy (ref. SG101803) and a large grant from the Arts and Humanities Research Council (ref. AH/K007084/1). I am indebted to Salih Akin and Ergin Öpengin for their help in organising the fieldwork, and to Hazel Gardner and Carlie Hanson for data archiving support. I am also grateful to the Editors for helpful comments on earlier versions.

References

Ahmed, Samer. 2016. *A grammar of Xerbi Kurmanji*. University of Manchester MA thesis.

Beavers, John. 2011. On affectedness. *Natural language and linguistic theory* 29(2). 335–370.

Fattah, Ismail Kamāndār. 2000. *Les dialectes kurdes méridionaux: Étude linguistique et dialectologique* (Acta Iranica 37). Leuven: Peeters.

Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A Construction Grammar approach*. Berlin: Mouton de Gruyter.

Haig, Geoffrey & Öpengin, Ergin. 2014. Kurdish: A critical research overview. *Kurdish studies* 2(2) [Special issue: *Kurdish linguistics: Variation and change*]. 99–122.

Haig, Geoffrey & Öpengin, Ergin. 2018. Kurmanji Kurdish in Turkey: Structure, varieties and status. In Bulut, C. (ed.), *Linguistic minorities in Turkey and Turkic-speaking minorities of the peripheries*, 157–229. Wiesbaden: Harrassowitz.

Jügel, Thomas. 2014. On the linguistic history of Kurdish. *Kurdish Studies* 2(2). 123–142.

MacKenzie, David Neil. 1961a. *Kurdish dialect studies*. Vol. 1. London: Oxford University Press.

MacKenzie, David Neil. 1961b. The origins of Kurdish. *Transactions of the Philological Society* 60(1). 68–86. [repr. with *Addenda and Corrigenda* in *Iranica diversa*, Roma: IsIAO, 369–387, 674–676].

Matras, Yaron. 1997. Clause combining, ergativity, and coreferent deletion in Kurmanji. *Studies in language* 21(3). 613–653.

Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.

Matras, Yaron. 2005. The classification of Romani dialects: A geographic-historical perspective. In Halwachs, D. & Schrammel, B. & Ambrosch, G. (eds.), *General and applied Romani linguistics*, 7–26. Munich: Lincom Europe.

Matras, Yaron et al. 2016. *The dialects of Kurdish*. Manchester: University of Manchester. (`http://kurdish.humanities.manchester.ac.uk/`) (Accessed 2019-05-31).

Matras, Yaron & Elšík, Viktor. 2008. *Romani Morpho-Syntax (RMS) Database*. (`http://romani.humanities.manchester.ac.uk/rms/`) (Accessed 2019-05-31).

Matras, Yaron & White, Christopher & Elšík, Viktor. 2009. The RMS Database and web resource. In Everaert, M. & Musgrave, S. (eds.), *Linguistic databases*, 329–362. Berlin: Mouton de Gruyter.

McCarus, Ernest N. 2009. Kurdish. In Windfuhr, G. (ed.), *The Iranian languages*, 587–633. London/New York: Routledge.

Öpengin, Ergin. 2016. *The Mukri variety of Central Kurdish: Grammar, texts, and lexicon*. Wiesbaden: Reichert.

Öpengin, Ergin & Haig, Geoffrey. 2014. Regional variation in Kurmanji: A preliminary classification of dialects. *Kurdish studies* 2(2). 143–176.

Thackston, Wheeler M. 2006. *Sorani Kurdish: A reference grammar with selected readings*. Cambridge, MA: University of Harvard. (`https://sites.fas.harvard.edu/~iranian/Sorani/sorani_complete.pdf`) (Accessed 2019-06-12).

11
Accounting for the combinations of clitic and affix person markers in Central Kurdish

Ergin Öpengin

Abstract: Central Kurdish possesses several sets of person markers, including the typical pronominal clitics and verbal agreement affixes of West Iranian languages. In past tense transitive constructions, the pronominal clitics mark the agent, while the verb agreement affixes mark the patient of the clause. In some particular constellations of past transitive constructions, these historical pronominal clitics and verb agreement suffixes have to occur in combination. The resulting ordering of clitic and verbal person markers shows a split: with all persons, except the third person singular, the pronominal clitic is first in combination. This ordering poses a theoretical problem: following the criteria of clitichood (Anderson 2005; Halpern 2001), clitics are expected to occur external to affixes, and the (not-clitic-like) idiosyncratic behaviour of the third person singular clitic requires separate treatment. Relying on Prosodic Phonology (Selkirk 1995) and through a closer look at the facts of lexical stress in Central Kurdish, I propose an analysis where the (historically affixal) verbal person markers in Central Kurdish are argued to be clitic-like in that they do not form a prosodic word with their host. Once this is established, the ordering of pronominal clitics before verbal person markers is argued to follow from a more general prosodically-defined second-position placement principle of pronominal clitics in Central Kurdish (Öpengin 2013). Finally, the idiosyncratic ordering of the third person singular pronominal clitic after verbal agreement affixes is explained in terms of constraints (Yip 1998) that require the forms in sequence to preserve their morpho-phonological and morpho-syntactic identity.

1 Introduction

Central Kurdish has a complex system of person markers (PM), including the historical pronominal clitics and verb agreement suffixes of West Iranian. In certain syntactic constellations, these historical pronominal clitics and verb agreement suffixes have to occur in combination. For instance, in a past tense transitive construction, the transitive subject is marked via a clitic PM, and the object is marked via a verbal PM. When in such a construction the bare verb is the first constituent in the verb phrase, the two person forms (clitic and verbal PMs) occur in sequential order. The resulting ordering of clitic and verbal PMs shows a split: with all persons, except the 3SG, the clitic marking the subject is first in combination, leading to a sequence as [HOST=CLITIC PM-VERBAL PM], as illustrated in (1). When the subject is 3SG, the clitic PM occurs second in combination, leading to a reversed sequence as [HOST-VERBAL PM=CLITIC PM], as illustrated in (2).

(1) *nārd-ū=yān-im* *bo ēre*
send.PST-PTCP=3PL-1SG to here
'They have sent me over to here.'

(2) *bird-īn=ī*
take.PST-1PL=3SG
'He took us.'

If, in terms of their morpho-phonological categorial status, historical clitic PMs are "clitics" and the verb agreement suffixes are "affixes", the general ordering of clitic and verbal PMs in such combinations, (as in (1) with a clitic preceding an affix) poses a theoretical problem: as syntactic elements, the clitics are expected to occur external to inflectional affixes. This is indeed one of the most reliable diagnostics of clitichood in the literature (Anderson 2005; Halpern 2001). On the other hand, the exceptional ordering of a 3SG clitic in combination with verbal PMs does conform to this theoretical expectation, but it is the exceptional or idiosyncratic ordering, and, as such, needs to be accounted for in its own respect.

These two separate but related questions concerning the ordering of clitic and affixal PMs in Central Kurdish have recently attracted the attention of the scholars working on Kurdish; however, a fully satisfactory analysis has not yet been proposed. Relying on the insights from Prosodic Phonology, I will propose an analysis where I will argue that the (historically affixal) verbal person markers in Central Kurdish are clitic-like in that they do not form

a prosodic word with their host. Once this fact is established, the occurrence of clitic PMs internal to "affixal" verbal PMs will no longer pose a theoretical problem, since the issue will then be reduced to the ordering of two clitic (or clitic-like) elements. The ordering of clitic PMs before verbal PMs will be argued to follow from a more general prosodically-defined second-position placement principle of clitic PMs in Central Kurdish (cf. Öpengin 2013). The idiosyncratic ordering of the 3SG clitic after verbal PMs (along with two other analogical constructions, see further below) is explained in terms of constraints that require the forms in sequence to preserve their morpho-phonological and morphosyntactic identity.

The data presented in this paper are naturalistic, spoken language data that were collected in the field in the speech zone of the Mukri dialect of Central Kurdish in North-west Iran (cf. Öpengin 2016 for extensive description of the fieldwork and corpus).

2 Forms and functions in argument indexing in Central Kurdish

Leaving aside the independent person forms, four sets of bound person markers are used for indexing clausal arguments in Central Kurdish. These are given in Table 1.

Table 1: Bound person marker paradigms in Central Kurdish

		Clitic PMs	Verbal Affix PMs		Copular PMs
			Set_1	Set_2	
SG	1	*=im*	*-im*	*-im*	*=im*
	2	*=it*	*-ī*	*-ī*	*=ī*
	3	*=ī*	*-ē*	*-Ø*	*=e*
PL	1	*=mān* / *=in*	*-īn*	*-īn*	*=īn*
	2	*=tān* / *=ū*	*-in*	*-in*	*=in*
	3	*=yān*	*-in*	*-in*	*=in*

These person forms are the two sets of verbal affix person markers, which derive from historical verb agreement suffixes and are used only with verb stems; a set of clitic person markers which derive from historical pronominal clitics of West Iranian (cf. Korn 2009) and can occur with hosts from diverse word categories; and finally, a set of copular person forms, which have most

probably evolved from the contraction of the present tense stem of the verb for 'to be' with verb agreement suffixes, whence their close formal similarity.

Reflecting the well-known split-ergativity of West Iranian (cf. Haig 2008), the person marker paradigms have different functions in present and past tense constructions. In present tense, the Set_1 verbal PMs are used for indexing the subject, while the clitics pronominally (i.e. non-obligatorily) mark the object and oblique arguments. In past tense, on the other hand, the clitics obligatorily mark the subject of transitive constructions (A), while the Set_2 verbal PMs mark (i) obligatorily the subject of intransitive constructions (S), and (ii) mostly pronominally mark the direct object[1] (P) and oblique arguments (R). Copular forms are used mainly with non-verbal predicates in present, and in some T-A-M forms of verbs based on past participle. Functional distribution of person markers can be summarized as in Figure 1.

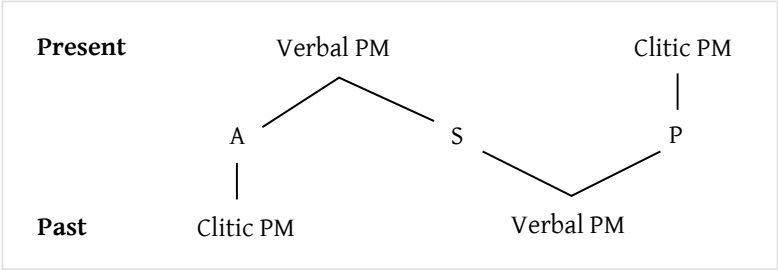

Figure 1: Distribution of person markers for marking clausal arguments (adapted from Bonami & Samvelian (2008))

Restricting our discussion to the past tense, the historically "pronominal" clitics have become obligatory "agreement" markers, since they occur even in the presence of their coreferent NP in the same clause, as in (3).

(3) *[qerewoł-ān]$_i$ kut=**yān**$_i$*
guard-PL say.PST=3PL
'The guards said [that ...]' (KF.132)

[1]Occasionally, the verbal affix person marker can be used even when the coreferent controller noun phrase is present in the clause, thus showing "agreement" with the logical object (P) of the clause. The facts behind this optional "agreement" are yet to be clarified, but the issue does not have any direct impact on our questions here.

The historical verb "agreement" suffixes have remained so in their function of marking an intransitive subject, but they have become "pronominal" when used for marking direct (4) and oblique objects (5).

(4) *hīč kes řā=ī-ne-de-girt-**m***
no person PVB=3SG-NEG.PST-IPFV-keep.PST-1SG
'No one would accept me in [their houses].' (ŽB.024)

(5) *xełik **lē**=yān de-kiřī-**m***
people from=3PL IPFV-buy.PST-1SG
'People would buy (the melons) from me.' (KF.021)

The comparison of (4) with (6) shows that the verbal PMs in these functions are used only when their coreferent NP is not present in the clause.

(6) *gödirēž-eke-ān=mān best-ewe*
donkey-DEF-PL=1PL tie.PRS-ASP
'(At the courtyard of the boy) we tied the donkeys.' (ŽB.047)

3 Clitic placement and sequential order of clitic and affix PMs in Central Kurdish

Although verbal affix PMs occur only with verb stems, clitics are mobile and can occur with any word category. They are placed in a prosodically defined "second" position within the Verb Phrase, occurring after (or within) the first stressed morphological or syntactic element in the VP (cf. Öpengin 2013: Ch. 5 on clitic placement principles in Central Kurdish). Accordingly, any of the following syntactic and morphological items that comes first in VP is a host for clitic PMs (restricted here to A-past clitics).

Placement of clitic PMs (within Verb Phrase)
NP > Nominal component of a complex predicate > Adposition > Preverb > pre-stem NEG/TAM > Verb stem (with inflection)

Now, when the bare stem (or with inflected verb) is the first/only item in VP, both the A-past clitic and the verbal PM indexing P or R will occur on the verb stem, leading to a combination of one clitic PM and one affix PM. Being syntactic elements, clitics would be expected to occur external to affixes

– the sole reliable criterion of clitichood according to some scholars (Anderson 2005; Halpern 2001; Siewierska 2004) – in the following order: [HOST-AFFIX=CLITIC].

However, in Central Kurdish, as it was illustrated with the sentence (1), in such combinations of clitic and verbal affix PMs, the clitic comes before the verbal PM (except with 3SG clitic), in the following order: [HOST=CLITIC-AFFIX]. The order in such combinations is given in paradigmatic form for the verb *girtin* 'to hold, to keep' in Table 2, with the 2PL being the invariant subject (A-past).

Table 2: Paradigm of the verb *girtin* 'to hold' with combinations of clitic and affix person markers

			A	P
SG	1	*girt*	***=tān***	***-im***
	2	*girt*	***=tān***	***N.A.***
	3	*girt*	***=tān***	***-Ø***
PL	1	*girt*	***=tān***	***-īn***
	2	*girt*	***=tān***	***N.A.***
	3	*girt*	***=tān***	***-in***

Furthermore, when the subject is 3SG, the order in combination is reversed, to become [HOST-AFFIX=CLITIC], with the clitic following verbal PM, as illustrated in (2). The ordering with a 3SG A-past clitic is thus exceptionally different from the ordering of the rest of the clitic PM paradigm, making it "idiosyncratic". The ordering of a 3SG clitic is given in paradigmatic form for the verb *birdin* 'to take' in Table 3 (the 2SG object is exceptionally realized as a clitic, with which we will deal separately further below).

Table 3: Paradigm of the verb *birdin* 'to take' with a third person singular pronominal clitic

			P	A
SG	1	*bird*	***-im***	***=ī***
	2	*bird*	***=it***	***=ī***
	3	*bird*	***-Ø***	***=ī***
PL	1	*bird*	***-īn***	***=ī***
	2	*bird*	***-in***	***=ī***
	3	*bird*	***-in***	***=ī***

There are thus two separate – but related – questions to resolve. First, given that the clitics are syntactic and affixes are morphological elements, we would expect the clitics to occur second in a clitic and affix combination, but in Central Kurdish the order turns out to be the reverse, where the clitic PMs come before the verbal affix PMs. Thus, the first question is how to approach this unexpected order in clitic and affix combinations. Second, although the order with a 3SG subject clitic conforms to the theoretical expectations (i.e. clitics occurring external to affixes), it turns out to be idiosyncratic within Central Kurdish morphosyntax, requiring an explanation of its own. Before presenting the analysis proposed here, we will have a quick look at the existing accounts of these problems in Kurdish linguistics.

3.1 Previous accounts of the problems

Considering the issue of the clitic person markers occurring between a stem and a verbal person marker, Samvelian (2007: 270–272) claims that the clitic PMs should be viewed as "affixes", since they occur inside a word and they "linearize" with respect to other morphological items, including lexical affixes (e.g. verbal person markers). The author sees further support for her analysis of clitic PMs as "affixes" in the idiosyncratic placement of 3SG clitics, since the clitics, unlike affixes, are not expected to show idiosyncratic behaviour (cf. Zwicky & Pullum 1983: 505). On that account, once it is accepted that the clitic PMs in Central Kurdish are "affixes", then there is no conceptual problem with the resulting ordering, since then they will be seen merely as instances of affix sequences. This account is not satisfactory, primarily since (i) clitic PMs show major clitic-like behaviour (e.g. varied host selection; non-phonological attachment to their host, etc.); (ii) the clitic person markers do not "linearize" with other affixes in the sense of inflection, but

rather they are placed following the morphological-syntactic context. Accordingly, the clitic occurs between the inflected stem and the verbal PM in (7a), but when, as in (7b), there is a host available earlier in the construction (the negative formative *ne-*), the clitic will be coaxed from its position and move to the earlier available position. This is not possible for typical inflectional categories.

(7) a. *nārd-ū=**yan**-im* → b. *ne=**yān**-nārd-ūw-im*
send.PST-PTCP=3SG-1SG NEG=3PL-send.PST-PTCP-1SG
'They have sent me.' 'They have not sent me.'

Finally, the analysis in Samvelian (2007) is built upon the assumption that no "second-position" (whether syntactic, morphological, or prosodic) can account for the placement of clitic PMs in Central Kurdish. However, as I have shown in Öpengin (2013: Ch. 5), the clitic person markers do indeed follow a prosodically-defined second position, which, to simplify, is following the first stressed element in the verb phrase (thus after or within the first prosodic word).

A second attempt, Haig (2008), proposes an analysis in terms of person-role constraint. In this account, the unmarked person-role constellations are when the Speech-Act-Participants (SAP; 1st/2nd persons) express an A argument, and non-SAPs (3rd persons) express a P argument. When this requirement is met, that is, the A is a SAP, the clitic occurs first in combination, acting "as a suffix" (Haig 2008: 293), whereas when the A-past clitic is a non-SAP, it follows – although the author accepts that with 3PL A-past, the clitic mostly precedes.

This account is more an attempt at explaining the variation in the order itself (A-P vs. P-A) than the unexpected clitic-affix ordering *per se*, since it does not, for instance, account for why assuming the A role would lead 1st and 2nd person clitics to behave "suffix-like" in the first place. Furthermore, it posits a dual nature for clitics, since they behave both like clitics (when following the verbal person markers) and like affixes (when preceding the verbal person markers). Finally, this account predicts the 3rd person clitics to occur second in combination; however, the 3PL clitic *=yān* systematically occurs first in combination in Mukri, while they mostly do so also in Suleimani dialect, on which the author's analysis is based. In the same vein, there is no person-role constraint in the ordering of a 3SG A-past and 3PL P-past, since both are non-SAP, but their ordering is the same as in the rest of the paradigm where the clitic follows the verbal PM (see the data in Table 3).

The existing accounts have greatly contributed to the elucidation of the practical and theoretical problems with clitic placement, and clitic and affix combinations in Central Kurdish. However, as it stands, we should still strive for a unified and satisfactory analysis of the remaining facts. In what follows, I propose an analysis of both sets of questions by looking more closely at the facts of prosody in Central Kurdish.

3.2 A prosodic analysis of the clitic and affix combinations

In my analysis below on the problems relating to the ordering of clitics in clitic/affix combinations in Central Kurdish, I argue that the verbal PMs are clitic-like, since they attach non-phonologically to their host. Once this is established, the occurrence of the clitic PMs between verb stem and verbal PMs follows from more general clitic placement principles as described above. We should then first look into the stress assignment rules in the language in order to demonstrate the non-phonological attachment of the verbal person markers.

The unmarked lexical stress is on the final syllable in Central Kurdish, as shown in (8) (the dots separate syllables and the sign (ˈ) indicates the syllable carrying the lexical stress):

(8) *hawīn* [ha.ˈwin] ‘summer’
bāyinjān [ba.jɨn.ˈʤan] ‘tomato’
gełāwird [gæ.ɬa.ˈwɨrd] ‘tiny leaf’

Regular inflectional affixes do not cause any change in this pattern, since they form a prosodic/phonological word with their host, as illustrated in Table 4 for definiteness (cf. MacKenzie 1961: 48), plural, participle suffixes:

Table 4: Assignment of lexical stress with affixes

Syllabic	Morphemic	Gloss	Translation
ti.rē.ye.ˈke	*tirē-eke*	grape-DEF	‘the grape’
kē.ˈlān	*kē-lān*	gravestone-PL	‘gravestones’
mir.ˈdū	*mird-ū*	die.PST-PTCP	‘dead’
kir.ˈdū.e	*kird-ūw-e*	do.PST-PTCP-COP.3SG	‘has done’

In the same vein, the verbal person markers of Set_1 follow this stress assignment pattern and receive word-final lexical stress, as shown in Table 5 with various person forms and verb stems:

Table 5: Assignment of lexical stress in verb forms with present tense verbal PMs

Syllabic	Morphemic	Gloss	Translation
dē.ˈnim	*de-hēn-**im***	IND-bring.PRS-1SG	'I shall bring (it).'
de.zā.ˈnī	*de-zān-**ī***	IND-know.PRS-2SG	'You know (it).'
de.zā.ˈnē	*de-zān-**ē***	IND-know.PRS-3SG	'S/he knows (it).'
de.ˈkeyn	*de-ke-**yn***	IND-do.PRS-1PL	'We shall do ...'
de.gi.ˈrin	*de-gir-**in***	IND-keep.PRS-3PL	'They keep ...'

Although formally almost identical with the Set_1 verbal PMs, the Set_2 verbal PMs differ from the latter by not receiving the unmarked word-final lexical stress of their host. They cause a change in the stress assignment pattern of the host verb to which they attach, such that the stress falls on the preceding (penultimate) syllable. This is visible in the intransitive and transitive verb forms in Table 6, where the stress systematically falls on the syllable immediately preceding the verbal PMs fulfilling various functions:

Table 6: Assignment of lexical stress in verb forms with past tense verbal PMs

Syllabic	Morphemic	Gloss	Translation
ˈnūs.tim	*nūst-**im***	sleep.PST-1SG	'I slept.'
gē.ˈřā.me.we	*gēřā-**m**-ewe*	relate.PST-1SG:R-ASP	'He narrated it to me.'
de.ˈgir.tī	*de-girt-**ī***	IPFV-keep.PST-2SG:R	'I would respect you.'
kir.ˈdū.wim	*kird-ūw-**im***	do.PST-PTCP-1SG:O	'You invited me.'

To restate the facts, in Central Kurdish, lexical stress is typically assigned to the last syllable of a phonological word. An affix attaching to a word would be expected to conform to this stress pattern, since it becomes part of the phonological word to which it attaches. This is the case for various well-known affixes in Central Kurdish (cf. Table 4 and Table 5), including the verbal person markers used with the present tense stem of verbs. However, the verb forms with verbal PMs of the Set_2 (i.e. those used with past tense verb stems) do not conform to this final-syllable stress assignment pattern. Instead, the stress pattern is altered such that the stress occurs on the syllable immediately preceding the verbal PM (cf. Table 6). What all these facts show

is that the Set_2 verbal PMs do not compose a phonological (or prosodic) word with their host, and as such, they are un-affix-like and more clitic-like in this most important respect.

Finally, there is highly relevant historical ground for such different behaviour of these person forms: The past tense verbal PMs (Set_2) most probably derive from the contraction of the verb stem *ha-* 'to be' and verb agreement suffixes. Compare the Middle Persian example in (9), where the past transitive verb construction is periphrastic, consisting of a past participle and the verb 'to be', which in its turn carries the verb agreement marker (glosses adapted):

(9) *'w=t'n dryst (q)yrd* ***hym*** *(or* ***hy-m****)*
and=2PL healthy do.PTCP COP.1SG
'and you have cured me.'
(MacKenzie 1999 [1979], cited in Haig 2008: 124)

Now that we have established the clitic-like behaviour of verbal person markers, the occurrence of the clitic person markers internal to verbal person markers no more poses a conceptual problem, since the resulting sequence involves two clitic or clitic-like elements. To the contrary, their placement follows from the more general placement principles of the clitic PMs in CK. As I have briefly noted above (and extensively discussed in Öpengin 2013: Ch. 5), the clitic PMs in CK occur after the first stressed element in the verb, which, in this case, is the verb stem. Apart from conforming to this prosodically-defined "second-position", the placement of clitic PMs before the verbal person markers reflects two other more general facts. First, it conforms to "argument hierarchy", as the index marking the higher argument (A > P) comes first in combination. Second, it reflects the divergent grammatical status of the two sets of person markers, namely that the clitic person markers as grammaticalised "agreement" markers come before the non-obligatory "pronominal" verbal person markers.

4 Idiosyncrasies with the order and form of person markers in combination

The analysis above suggested that the verbal PMs are essentially clitics (or clitic-like), as they are only non-phonologically attached to their host verbs. If this is the case, the question still remains as to why, in past transitive constructions, the ordering of a 3SG subject clitic and object verbal PMs is excep-

tionally reversed to become one in which the clitic PM follows the verbal PM, hence violating the "second-position" clitic placement principle. This and an additional two analogical processes are dealt with in this section.

4.1 Idiosyncratic ordering with a 3SG subject clitic

As stated above, the regular order of clitic and verbal person markers is clitic PM preceding verbal PM, as in Table 7 – the paradigm of the verb *birdin* 'to take' conjugated with 3PL A-past clitic and all verbal person markers coding P:

Table 7: Regular order in clitic and verbal PM combinations (*birdin* 'to take' with 3PL A-past clitic)

			A	P
SG	1	*bird*	***=yān***	***-im***
	2	*bird*	***=yān***	***=it***
	3	*bird*	***=yān***	***-Ø***
PL	1	*bird*	***=yān***	***-īn***
	2	*bird*	***=yān***	***-in***
	3	*bird*	***=yān***	***-in***

However, with a 3SG subject, this order is reversed to one in which the clitic PM follows the verbal PM, as in Table 8 – the paradigm of the verb *birdin* 'to take' conjugated with 3SG A-past clitic and all verbal person markers coding P:

Table 8: Exceptional (reverse) order in clitic and verbal PM combinations with a 3SG A-past clitic (*birdin* 'to take')

			P	A
SG	1	*bird*	***-im***	***=ī***
	2	*bird*	***=it***	***=ī***
	3	*bird*	***-Ø***	***=ī***
PL	1	*bird*	***-īn***	***=ī***
	2	*bird*	***-in***	***=ī***
	3	*bird*	***-in***	***=ī***

I argue that this order of the forms in combination is exceptionally reversed due to OCP[2]-like phonological constraints that require the elements in a combination to be distinct (Yip 1998). That is, the paradigmatically expected ordering of the clitic PMs before verbal PMs blurs the morpho-phonological identity of the forms in sequence, thereby blocking the expression of the morphosyntactic information (i.e. roles of clausal arguments) encoded by those forms.

This blurring of the forms occurs principally because of a reduction in the number of syllables in the combinations of person markers. Notice that the regular clitic and verbal PM combinations are systematically bi-syllabic (except with the zero-form 3SG verbal PM), cf. Table 2 and Table 7. The idiosyncratic ordering with 3SG A-past clitic (Table 8) likewise results in sequences consisting of bi-syllabic units. However, the hypothetical "regular" or "paradigmatically expected" order with the 3SG A-past clitic preceding the verbal PM, as in Table 9, systematically leads to sequences of person forms that are chunked into mono-syllabic units.

Table 9: Expected "regular" order with a 3SG A-past clitic

			A	P
SG	1	*bird	**=ī**	**-im**
	2	*bird	**=ī**	**-y**
	3	*bird	**=ī**	**-Ø**
PL	1	*bird	**=ī**	**-īn**
	2	*bird	**=ī**	**-in**
	3	*bird	**=ī**	**-in**

The 3SG clitic PM is the only member of the paradigm of clitic PMs which does not contain a consonant in its form. Thus, when it is combined with the verbal PMs in the putative expected manner [HOST=CLITIC PM-VERBAL PM], the distinct formal identity of the forms (especially that of the clitic PM) is lost. For a better illustration, the form sequences resulting from the expected hypothetical and actual (idiosyncratic) orders are given in Table 10.

As it can be noted in Table 10, the expected clitic-affix sequences do not conveniently reveal the morpho-phonological identity of the forms in combination. This is so because, on the one hand, the two forms in combination are merged into mono-syllabic units, and, on the other hand, the resulting

[2]Obligatory Contour Principle

form sequences are either identical or similar to single person forms (such as *īn*, which is identical with 1PL verbal PM). Thus, given that the person forms in combination are unidentifiable as distinct elements, they cannot express the morphosyntactic information of A and P roles, respectively, that they encode in the input. To avoid such an inefficient construction ("identity avoidance" in the sense of Yip 1998), the order of the respective person forms is swapped, with the clitic following the affix, as in Table 8. The combinations obtained through this paradigmatically irregular ordering transparently reveal the morpho-phonological identity of the person forms in combinations (except for the combination with 2SG verbal PM, which results in another exceptional outcome, see further below), as it can also be seen in Table 10 and Table 8. That is, the elements in the combinations are still distinct, decomposable into definite individual person forms, and thus fully capable of expressing the morphosyntactic information they carry in the clause.

Table 10: Morpho-phonological form of clitic and affix combinations

Expected clitic-affix order	Actual affix-clitic order
*=*īm*	*-imī*
*=*īy*	*-*īy*
*=*ī*	*-ī*
*=*īn*	*-īnī*
*=*īn*	*-inī*
*=*īn*	*-inī*

In short, there are principally two crucial and interrelated conditions that are not met in the expected combinations of the 3SG clitic PM with verbal PMs. These are morphological and phonological distinctiveness of the person forms in combination, and the identifiability of the morphosyntactic information expressed by the person forms. Both of these conditions are met by swapping the positions of the person forms such that the clitic PM follows the verbal affix PM in combination. This violates the "second-position" clitic placement principle working in the rest of the clitic PM paradigm. However, in this instance, the distinctive and unambiguous expression of the person forms in question – and thus the transfer of the morphosyntactic information they encode – ranks higher than the clitic placement principles.[3]

[3] For an account in terms of ranking hierarchies within Optimality Theory for these and similar problems in the morphosyntax of Central Kurdish, see Öpengin (2013: Ch. 6).

4.2 Idionsycratic ordering of 3SG clitic and 3SG copular PMs

In parallel to their interaction with verbal PMs, clitic PMs may occur in combination with copular personal endings in a number of constructions, such as present perfect tense (10) and predicative possession (11). With all persons except a 3SG clitic PM, the order within combination is [HOST=CLITIC PM-COPULAR PM], as illustrated in (10a) and (11a). However, when a 3SG clitic PM enters into play, whether in A-past function (10b) or as the possessor (11b), the order is reversed such that the copular PM comes before the clitic PM within the sequence.

(10) a. *ew xezīne-ī dizī-w=**mān-e***
DEM treasure-EZ steal.PST-PTCP=1PL-COP.3SG
'(This treasure that) we have stolen' (MK.298)

b. *nūsī-w-**yet=ī***
write.PST-PTCP-COP.3SG=3SG
'He has written (as: ...)' (MK.143)

(11) a. *le ew pārče-ān-e-ī he=**mān-e***
from DEM fabric-PL-DEM1-EZ exist=1PL-COP.3SG
'[...] (out of these fabrics that) we have' (HF.062)

b. *dū ser meř-e-y ke he-**yet=ī***
two CLASS sheep-DEM1-EZ REL exist-COP.3SG=3SG
'the two heads of sheep that he possesses' (FN)

This reversal in the order of person forms is likewise morpho-phonologically motivated: The regular and expected combination of a 3SG clitic PM with a 3SG copular ending would be *=ī-(y)e*, with the clitic coming first within the sequence. This would systematically change the vowel <ī> ([i]) to a glide <y> ([j]), giving the sequence *=ye* (analysed as *=y-e*) with the clitic first and the copular ending second. However, the phonological form of the sequence, on the one hand, does not preserve the distinctive input form of the clitic PM (i.e. <ī>) and, on the other hand, is principally identical with the form of the 3SG copular ending *-(y)e*. The reversal of the order of the person forms steps in to avoid identity. That is, the reversal of the order of the person forms in sequence resituates the distinctive form of the clitic PM as *=ī*. This is enabled by the resurfacing of a (historical) final <t> (preceding a vocalic element) in the form of the 3SG copular ending, giving a distinctive *-(y)et* form

to it, which in combination with a 3SG clitic yields the complex form *-(y)etī*. (The resulting *-(y)etī* form is not usually correctly analysed by native speakers and native grammarians as consisting of a copular ending plus a clitic PM. However, in light of the motivation provided here and observed elsewhere in the morphosyntax of the language, its analysis and decomposition here is rather straightforward.)

4.3 Disformation of 2SG verbal PM into clitic PM

It was seen in the discussion above, for instance in Table 8, that the combination of a 3SG subject clitic with 2SG verbal PM expressing the object results in a change in the form of the latter (though not in all dialects). That is, a sequence of 2SG verbal PM *-ī* and 3SG clitic PM *=ī* would normally give *-īy* (analysed as: *-ī=ī*), as in (12a). However, as we can see in (12b), and in Table 8, the resulting sequence is *=itī*, analysed as *=it=ī*, namely the sequence of a 2SG clitic PM and a 3SG clitic PM.

(12) a. EXPECTED VERB FORM

bird-ī=ī *(> birdīy)*
take.PST-2SG=3SG

'S/he took you.'

b. OBSERVED VERB FORM

bird=it=ī
take.PST=2SG=3SG

'S/he took you.'

In this case, the 2SG verbal PM *-ī* is "disformed" into the corresponding 2SG clitic PM *=it*. This formal change is again motivated by a concern to avoid identity and keep the two input person forms morpho-phonologically distinct. While the expected combination *–īy* (/ij/), consisting of two vocalic dorso-palatal segments[4], may not straightforwardly achieve this task, the sequence resulting from "disformation" of the verbal PM into its corresponding clitic PM quite unambiguously achieves it.

In this section, I have discussed three instances of exceptional or paradigmatically irregular orderings and person form realizations. These were the reverse ordering of a third person singular clitic with verbal person markers and copular personal endings, and the disformation (or "suppletion") of the second person singular verbal person marker into corresponding clitic person marker. All three phenomena were shown to be motivated by a concern

[4]Note that, in some circles, for their features the glides are considered to be identical with their vowel counterparts. The distinctions are considered to be at the phonetic level. See, for instance, Padgett (2008).

to keep the morpho-phonological distinctiveness of the person forms when they form a sequence or combination. Central Kurdish data is by no means unique in this respect. A number of parallel phonological and morphological constraints on the organization of clitic sequences have been documented in Gerlach (2002), where the author shows, for instance, that phonologically similar elements may result in the dissimilation, suppletion, or deletion of a clitic within clitic combinations.

5 Summary and conclusion

In this paper, I have dealt with a number of theoretical and practical questions concerning the concatenation of person markers from different categories in the past tense transitive constructions in Central Kurdish. These questions were (i) the ordering of clitic person markers before verbal (affix) person markers, (ii) the idiosyncratic placement of a third person singular clitic after verbal and copular person markers – which otherwise come before the clitic person marker – and, finally, (iii) the disformation of a second person singular verbal person marker into a corresponding second person singular clitic person marker.

The first problem was accounted for by postulating that the verbal person markers, although historically deriving from verb agreement suffixes, are synchronically clitic-like in that they only non-phonologically attach to their host verb. Evidence for this analysis came from the facts of stress assignment in the language: The unmarked stress pattern in the language is on final syllable. The affixes of diverse sort were illustrated to conform to this pattern when attaching to their host words. However, the verbal person markers were shown to cause a change in the final-syllable stress pattern by ousting the stress onto the immediately preceding syllable. This proved that the verbal person markers, unlike typical affixes, only non-phonologically attach to their host verb, and, as such, are clitics or clitic-like elements in this respect. Once this was established, the placement of clitic person markers in between the verb stem and the verbal person markers follows from more general clitic placement principles according to which the latter occur after the first stressed element in the verb (e.g. a prefix, preverb, or verb stem).

The paradigmatically exceptional placement of a third person singular clitic person marker with verbal person markers and copular personal endings was accounted for by constraints that required the forms in sequence to preserve their morpho-phonological identity, which, in turn, is necessary for the ex-

pression of the encoded morpho-syntactic information. Finally, the disformation of a second person singular verbal personal marker to a corresponding clitic person marker, when followed by a third person singular clitic, was again shown to be motivated by "identity avoidance", that is, to avoid a sequence of two homophonous person forms, which would fail to keep the morpho-phonological idendity of the person forms and, accordingly, fail to express the encoded morpho-syntactic piece of information.

The above discussion reveals two important aspects of more general importance. First, a closer and holistic look into the language material (e.g. prosody) may provide a more straightforward and, hopefully, convincing answer to apparent problems. Second, although the categories – such as clitic and affixes – have prototypical properties that are predictive of certain types of behaviour, the categorial restrictions can be – and often are – overridden when higher-ranking requirements of efficient expression of encoded morpho-syntactic information are at stake.

References

Anderson, Stephen R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.

Bonami, Olivier & Samvelian, Pollet. 2008. Sorani Kurdish person markers and the typology of agreement [presentation handout]. *13th International Morphology Meeting*, Vienna, 03 February 2008.

Gerlach, Birgit. 2002. *Clitics between syntax and lexicon*. Amsterdam/Philadelphia: John Benjamins.

Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A Construction Grammar approach*. Berlin: Mouton de Gruyter.

Halpern, Aaron. 2001. Clitics. In Spencer, A. & Zwicky, A. M. (eds.), *The handbook of morphology*, 101–122. Oxford: Blackwell.

Korn, Agnes. 2009. Western Iranian pronominal clitics. *Orientalia suecana LVIII*. 159–171.

MacKenzie, David N. 1999 [1979]. Mani's Šābuhragān. In Cereti, C. G. & Paul, L. (eds.), *Iranica diversa*, vol. 1, 83–117. Rome: Instituto Italiano per L'Africa e l'Oriente.

MacKenzie, David Neil. 1961. *Kurdish dialect studies*. Vol. 1. London: Oxford University Press.

Öpengin, Ergin. 2013. *Clitic/affix interactions: A corpus-based study of person marking in the Mukri variety of Central Kurdish*. University of Sorbonne-Nouvelle & University of Bamberg dissertation.

Öpengin, Ergin. 2016. *The Mukri variety of Central Kurdish: Grammar, texts, and lexicon*. Wiesbaden: Reichert.

Padgett, Jaye. 2008. Glides, vowels and features. *Lingua* 118(12). 1937–1955.

Samvelian, Pollet. 2007. What Sorani Kurdish absolute prepositions tell us about cliticization. In Hoyt, F. & Seifert, N. & Teodorescu, A. & White, J. (eds.), *Texas Linguistic Society IX: The morphosyntax of underrepresented languages*, 263–283. Stanford: CSLI Publications.

Selkirk, Elisabeth. 1995. The prosodic structure of function words. In Beckman, J. & Walsh Dickey, L. & Urbanczyk, S. (eds.), *Papers in Optimality Theory*, 439–469. Amherst, MA: Dept. of Linguistics, University of Massachusetts.

Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.

Yip, Moira. 1998. Identity avoidance in phonology and morphology. In Lapointe, S. G. & Brentari, D. K. & Farrel, P. M. (eds.), *Morphology and its relation to syntax and phonology*, 216–247. Stanford: CSLI Publications.

Zwicky, Arnold M. & Pullum, Geoffrey K. 1983. Cliticization vs. inflection: English *n't*. *Language* 59(3). 502–513.

University
of Bamberg
Press

Current issues in Kurdish linguistics contains ten contributions which span the field of Kurdish linguistics, both in terms of geography and in terms of the range of topics. Along with several works on Kurmanji (Northern Kurdish) and Sorani (Central Kurdish), two chapters shed light on the lesser-known Southern Kurdish language area. Other studies are comparative, and treat the Kurdish language area in its entirety. The linguistic approaches of the authors are a mix of formal and typological perspectives, and cover topics ranging from geographical distribution and variation to phonology, morphosyntax, discourse structure, historical morphology, and sociolinguistics.

The present volume is the first of its kind in bringing together contributions from a relatively large number of linguists, working in a diverse range of frameworks and on different aspects and varieties of Kurdish. As such, it attests to the increasing breadth and sophistication now evident in Kurdish linguistics, and is a worthy launch for the new series Bamberg Studies in Kurdish Linguistics (BSKL).

ISBN: 978-3-86309-687-8

9 783863 096878

www.uni-bamberg.de/ubp