10.20378/irb-56764

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).