Creating an Electronic Lexicon for the Under-resourced Southern Varieties of Kurdish Language

Zahra Azin[1], Sina Ahmadi[2]

[1]Geomatics and Cartographic Research Center, Carleton University
[2]National University of Ireland Galway, Ireland
E-mail: zahraazin@cmail.carleton.ca, ahmadi.sina@outlook.com

Abstract

Thanks to the advances in information technology and communication, many endangered, vulnerable and under-represented language communities have a chance to revitalize and document their languages. In comparison to other Kurdish variants such as central Kurdish (also known as Sorani) and northern Kurdish (also known as Kurmanji), southern Kurdish has received trivial attention making it an under-documented and under-resourced language that is spoken primarily in the Kurdish regions of Iran, particularly Kermanshah and Ilam provinces. As the case of our study, we focus on creating an electronic monolingual lexicon of significant size for the southern variants of Kurdish in the Ontolex-Lemon ontology by converting a bilingual and monolingual dictionary. In addition, we report our efforts in using a semi-automatic pivot-based translation inference approach to align the current resource with other resources in Kurdish and Gorani languages. We believe that this resource increases inter-operability across various natural language processing systems and facilitates many tasks in computational linguistics for Kurdish. Our resource is publicly available under a Creative Commons Attribution-ShareAlike 4.0 International License[1].

Keywords: southern Kurdish; electronic lexicography; less-resourced languages; machine-readable dictionary

1. Introduction

Given the increasing importance of information technology and accessibility in our era, language communities around the globe are experiencing a momentous period to consolidate their languages with technology. As an initial step in documenting and processing natural languages, electronic resources, particularly lexicons, are of significance to pave the way for gradual and more advanced progress. That being said, many endangered and under-documented languages undergo further challenges due to the scarcity of language and linguistic resources.

In this paper, we focus on one of the under-represented variants of the Kurdish language, southern Kurdish. Kurdish language is an Indo-European language spoken by 20-30 million in the Kurdish regions of Iraq, Iran, Turkey and Syria, and also among the Kurdish diaspora around the world. Generally, the language is categorized as a less-resourced language with few linguistic resources and sparse documentation (Abdulrahman et al., 2019). Among the three main dialects of Kurdish, namely northern Kurdish or Kurmanji, central Kurdish or Sorani and southern Kurdish, the latter lacks resources comparatively more than the other two variants (Ahmadi, 2020). To remedy this, in this paper we discuss our efforts in creating an electronic lexicon for southern Kurdish.

Aware of the advances in the Semantic Web and Linked Data technologies, we focus on converting a printed dictionary, which is provided to us by a native lexicographer, into the Ontolex-Lemon ontology (McCrae et al., 2017). The dictionary is compiled based on lemmata of southern Kurdish and provides translations in Persian and Sorani Kurdish. In this regards, our methodology is based on Ahmadi et al. (2019)'s where the printed dictionary is semi-automatically converted into Ontolex-Lemon (McCrae et al., 2017). Ontolex-Lemon aims at modelling existing lexicographic resources as linked data and

---

[1] https://github.com/sinaahmadi/SKurdishLexicon

providing a conceptual model of language and linguistic objects to increase re-usability of lexicographic content by following Semantic Web standards. Thanks to the current advances in linguistic linked open data (LLOD), lexicographic resources are widely in use in Ontolex-Lemon nowadays; one compelling example is Wikidata[2] which openly provides access to data regarding lexemes, senses and lexical forms (Nielsen, 2020).

Moreover, as a preliminary study, we carry out a translation inference task where our southern Kurdish lexicon is aligned with the Sorani dictionary of (Ahmadi et al., 2019) at sense-level. In addition to lexicons which are crucial resources in many natural language processing (NLP) tasks such as word-sense disambiguation and spelling error correction, alignment of lexical resources has proved to be beneficial in many natural language processing tasks (Ahmadi et al., 2020).

2. Southern Kurdish

There are different approaches proposed by linguists and dialectologists to classify Kurdish. All these classifications contain a group representing a bundle of familiar varieties including Kalhori, Feyli, Kermashani, and Laki. In the classification of Kurdish varieties, Hassanpour (1992) names this group Kermashani and identifies it as one of the main varieties of Kurdish alongside Kurmanji, Sorani, and Hawrami. At the same time, (Izady, 1992: p. 169) identifies two main groups of Kurdish language namely Kurmanji which includes north Kurmanji and south Kurmanji which refers to Sorani, and Pahlawani which consists of Zaza and Gorani (also written as Gurani). According to this classification, dialects spoken in the southern areas of Kurdish speaking settlements (starting from eastern Turkey to western Iran) are considered as varieties of Gorani. Later, Fattah (2000) provided a clearer picture of Kurdish language based on a detailed fieldwork proposing a plausible classification of Kurdish into five groups of northern Kurdish (or Kurmanji), central Kurdish (or Sorani), southern Kurdish, Zazaki, and Hawrami (also referred to as Gorani). In this section, partly complying with Fattah's classification (Fattah, 2000), we discuss southern Kurdish (also called SK) and some of its issues.

Southern Kurdish is a variety of Kurdish language consisting of a group of vernaculars spoken by almost three million people across an extensive region of western Iran including Ilam, a large area of Kermanshah, and some parts of Lorestan and Kurdistan provinces (Fattah, 2000). As shown in Figure 1, this variety is also spoken in eastern Iraq in Khanaqin and Mandali, very close to the borders with Iran. Due to the geography of areas where southern Kurdish varieties are spoken, the population of southern Kurdish speakers is quite dispersed. On the other hand, as it is illustrated in the map, the presence of other languages such as Lori, mainly spoken in Lorestan, Chaharmahal and Bakhtiari, and parts of Ilam has resulted in a linguistic continuum between Kurdish and Lori in those areas (Aliakbari et al., 2015).

The existence of other languages and varieties such as Gorani, Lori, Persian, Turkic, and Arabic has resulted in a complex linguistic situation, language contact and multilingualism slowing down the progress of studies on southern Kurdish and its vernaculars. This complex linguistic situation poses challenges to the classification of southern Kurdish dialects. In such linguistic context, ethnic affiliation directly affects the categorization of language varieties. Therefore, the study of southern Kurdish is closely related to

ethnogeography which means the names of the vernaculars refer to specific ethnic groups or villages where their speaker reside. For instance, the names of some major variants such as "Kalhor" or "Kordali" have been taken from large tribes whereas variants with a smaller group of speakers are named after the geographical district such as "Malikshahi" spoken in a district called Malikshah. The variations of southern Kurdish are mainly mutually intelligible, however, as the geographical distance between the speakers increases, more effort is required to understand other varieties (Fattah, 2000).

Nevertheless, the language shift towards Persian among the southern Kurdish speakers due to sociolinguistic factors has been considered a threat to the native language of generations since the past few decades (Yarahmadi, 2021). This language shift, according to (Yarahmadi, 2021), not only has changed the vocabulary of the language in such a way that Kurdish words are replaced by Persian equivalents, but also the syntax and phonology of the language has undergone many changes.

Figure 1: Revised map of the distribution of southern Kurdish dialects (Fattah, 2000) from (Belelli, 2019: p. 3)

2.1 Dialects

Classification of southern Kurdish varieties is not easy mainly due to the lack of descriptive studies regarding the nuances among them. Fattah (2000) was the first one who outlined an initial classification of southern Kurdish vernaculars. He identified 27 sub-groups of southern Kurdish in Iran and eight in Iraq based on which he proposed seven main dialects including Bijari (also known as Garusi), Kolya'i (called Chardawri in Kurdistan Province of Iran), Laki spoken in the city of Kermanshah[3], Kalhori (including Sanjabi and Zangana), Malekshahi, Badre'i, and Kordali (Belelli, 2019).

---

[3] Whether Laki is a variety of Kurdish or Lori is still an open question and we avoid this discussion as it is not in the scope of the present study.

There are different factors affecting dialectal variations including geographic, social, and individual properties. The linguistic complexity of the area where southern Kurdish is spoken makes the study of varieties even more challenging due to an extensive population mobility, language contact, and complexity of intersecting some dialects. To represent a clear picture of this dialect and its varieties, an extensive fieldwork is required to be undertaken to better understand the linguistic characteristics of its dialects.

2.2 Scripts

It is not until recently that southern Kurdish has been used in writing, except for some literary and religious works. Historically, literary Gorani was the primary means for writing literary works in southern Kurdish (Kreyenbroek & Chamanara, 2013). Unlike other Kurdish varieties such as Sorani, southern Kurdish has never gone through language standardization mainly because of the dominance of official languages in regions where it is spoken. The speakers mainly use those languages such as Persian, Arabic, and Sorani Kurdish for writing. Today, southern Kurdish varieties are rarely found in written form or they simply follow other existing scripts. It seems that using such writing systems does not prevent this variety from expressing itself properly. However, some phonological features which distinguish southern Kurdish from other varieties have not been represented in such writing systems. Table 1 represents Kurdish scripts[4] and tentatively illustrates the place of southern Kurdish in this system. As the variety have not been standardized, different existing forms used in existing resources are shown in the table (Fattah, 2000; Jalilian, 2006). The examples in this manuscript are provided in the Latin-based script of Kurmanji Kurdish.

The phonological variations among southern Kurdish dialects depend on the region they are spoken in. The pharyngeal consonants [ħ], [ɣ] and [ʕ] are absent in southern Kurdish varieties spoken in Iran, unlike Sorani Kurdish. The voiced velar nasal [ŋ] seems to be missing in variations spoken in Khanaqin of Iraq and Qasr-e Shirin in Iran; however, it is common in Kalhori and Mandali dialects as in *řeŋ* 'color' (Fattah, 2000). In some Kalhori dialects (and along the border with Iraq) [gʲ] replaces word final [g] or [k], e.g. *seg^j* 'dog' (Belelli, 2019).

Unlike the consonants, vowels in southern Kurdish vary extensively among its dialects in different regions. Among vowels represented in Table 1, [ə] and [ü] are southern Kurdish specific and do not appear in other Kurdish varieties. The mid central /ə/ is used as Ezafe (also known as Izafe) marker in southern Kurdish to distinguish between /î/ in Sorani and /i/ in its varieties, e.g. *kuř-i xas* "(the) good boy" vs. Sorani *kuř-î xas* "a good boy" (Karimpour, 2003).

One of the main challenges before southern Kurdish standardization is phonological variations among its dialects. Language contact and multilingualism lead to gradual changes in the phonological system which, in turn, lead to a more complex linguistic situation in regions where this dialect is spoken. However, as the first step towards the standardization of this variety, a close investigation of the scripts used in existing resources is indispensable.

---

[4] In this table, following the Unified Kurdish Alphabet introduced by the Kurdish Academy of Language, the Latin based Yekgirtú has been used to represent Kurdish script. See http://www.kurdishacademy.org

Kurdish Phonemes (IPA)	Latin-based	Yekgirtû	SK (existing resources)	Arabo-Persian			
				initial	middle	final	single
[ɑː]	A a	A a	A a	ئا-	ـا-	ـا	ا
[b]	B b	B b	B b	بـ	ـبـ	ـب	ب
[ʧ]	Ç ç	C c	Ç ç Č č Ch ch	چـ	ـچـ	ـچ	چ
[dʒ]	C c	J j	C c J j Ĵ ĵ	جـ	ـجـ	ـج	ج
[d]	D d	D d	D d	دـ	ـدـ	ـد	د
[æ]	E e	E e	E e A a	ئه	ـهـ	ـه	ه
[eː]	Ê ê	É é	Ê ê	ئێـ	ـێـ	ـێ	ێ
[f]	F f	F f	F f	فـ	ـفـ	ـف	ف
[g]	G g	G g	G g	گـ	ـگـ	ـگ	گ
[h]	H h	H h	H h	هـ	ـهـ	ـه	ە
[ħ]	H h	H' h'	H h H' h'	حـ	ـحـ	ـح	ح
[ə]	I i	I i	I i				
[iː]	Î î	Í í	Î î	ئیـ	ـیـ	ـی	ى
[ʒ]	J j	Jh jh	J j Ž ž	ژ	ـژ-	ـژ	ژ
[k]	K k	K k	K k	کـ	ـکـ	ـک	ک
[l]	L l	L l	L l	لـ	ـلـ	ـل	ل
[ɫ]	Ł ł	Ll ll	Ļ ļ Ł ł Lˆ lˆ	ڵـ	ـڵـ	ـڵ	ڵ
[m]	M m	M m	M m	مـ	ـمـ	ـم	م
[n]	N n	N n	N n	نـ	ـنـ	ـن	ن
[oː]	O o	O o	O o	ئۆ	ـۆ-	ـۆ	ۆ
[p]	P p	P p	P p	پـ	ـپـ	ـپ	پ
[q]	Q q	Q q	Q q	قـ	ـقـ	ـق	ق
[ɾ]	R r	R r	R r	ر	ـر-	ـر	ر
[r]	Ř ř	Rr rr	Ř ř Ŗ ŗ	ڕ	ـڕ-	ـڕ	ڕ
[s]	S s	S s	S s	سـ	ـسـ	ـس	س
[ʃ]	Ş ş	Sh sh	Ş ş Š š Sh sh	شـ	ـشـ	ـش	ش
[t]	T t	T t	T t	تـ	ـتـ	ـت	ت
[ʊ]	U u	U u	U u	ئو	ـو-	ـو	و
[uː]	Û û	Ú ú	Û û	ئوو	ـوو-	ـوو	وو
[v]	V v	V v	V v	ڤـ	ـڤـ	ـڤ	ڤ
[w]	W w	W w	W w	و	ـو-	ـو	و
[x]	X x	X x	X x	خـ	ـخـ	ـخ	خ
[j]	Y y	Y y	Y y	یـ	ـیـ	ـیـ	ى
[z]	Z z	Z z	Z z	ز	ـز-	ـز	ز
[ʔ]	‘		‘	ئـ	-	-	-
[ʕ]		'	É é	عـ	ـعـ	ـع	ع
[ɣ]	X x	X' x'	X' x'	غـ	ـغـ	ـغ	غ
[ŋ]	ŋ		ŋ	-	ـنگـ	ـنگ	نگ
[gʲ]			Ğ ğ	گـ	ـگـ	ـگ	گ
[y]	Ü ü		Ü ü	ۆ	ـۆ-	ـۆ	ۆ

Table 1: Current forms of Kurdish alphabets found in existing resources

.

2.3 Vocabulary

Lexical differences in southern Kurdish varieties are found in dictionaries and everyday conversation of the speakers. However, using various lexical items does not obscure the intelligibility of the dialects for listeners who speak different varieties. Moreover, neighboring dialects and inter-dialectal means of communication result in the gradual change of the lexicon (Belelli, 2019). Speakers try to approximate their dialects to the highest prestige one and sometimes they do this to build solidarity across their differences.

Despite all those similarities, there still exist nuances in the lexical items used by the speakers of southern Kurdish dialects. As the provincial border of south-west Kermanshah is crossed to Ilam and eastern Iraq, lexical differences become more salient. The difference might be in the form of a simple shift of the vowels (e.g. *çö* vs. *çaw* 'eye'), or by using different words for a same concept (e.g. *keřemye* vs. *tem* 'fog').

One way to study dialect variation is the lexicon, or the vocabulary used by the speakers. Varieties might either use different words or same words with different meanings for instance, Badre'i speakers use *xwazî* 'want' to ask for something, while in Kalhori, speakers use the same verb when proposing a girl to get married. Although such differences might lead to misunderstanding, it does not interrupt the communication.

3. Approach

3.1 Dictionary Compilation

In this study, we use "Ferhengî Başur" (literally meaning "South dictionary"), a southern Kurdish-central Kurdish-Persian dictionary compiled and edited by Jalilian (2006) with the purpose of codification of southern Kurdish. Initially, the dictionary was created to be a part of "Henbane Borîne", a Kurdish-Kurdish-Persian dictionary written by Abdurrahman Sharafkandi known as Hazhar in 1990 (Sharafkandi, 1991). "Henbane Borîne" contains around 60,000 entries with lexemes from different Kurdish varieties. In addition, the Persian equivalent of entries along with a few examples are provided. "Ferhengî Başur" maintains the same structure but with a focus on the southern Kurdish varieties, particularly Kalhori and Laki.

Despite the attempt to document the general and folkloric vocabulary of southern variants of Kurdish in this resource, there is a lack of coverage of topics due to the scarcity of terminologies for Kurdish in general, and for these variants in particular. Similar to the majority of Kurdish dictionaries, our resource lacks consistent definition of entries in such a way that for only a few lemmata sense glosses are provided. The same issue can be observed with respect to idioms, examples and pronunciation. Among the words in various varieties of Kurdish, only those in Laki are specified by the lexicographer. Figure 2 illustrates the entry *qirtan* 'to cut' in southern Kurdish in the printed dictionary.
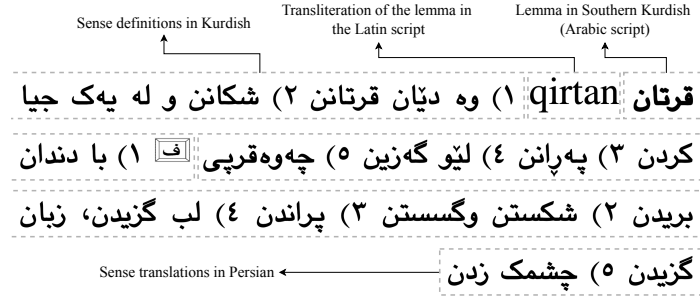
Figure 2: "*qirtan*" 'to cut', an entry in the printed version of Jalilian (2006: p. 534)'s Southern Kurdish dictionary

3.2 Conversion into Ontolex-Lemon

Ahmadi et al. (2019) propose an approach to create electronic lexicons for Sorani Kurdish, Kurmanji Kurdish and Gorani. Following the same approach, we use a semi-automatic technique to extract entries from the printed dictionary using regular expression. The extracted information is followed by a manual verification regarding the lemma in the Arabic-Persian script, its transliteration in the Latin script of Kurdish, glosses in Kurdish and their translations in Persian. In addition, a few entries are provided with additional information such as the sub-dialect where the word is used which are similarly included in the conversion process.

In order to increase the interoperability and accessibility of this resource, we provide the electronic dictionary in Ontolex-Lemon in the Resource Description Framework (RDF). The OntoLex-Lemon standard provides rich linguistic grounding for ontologies, such as representation of morphological and syntactic properties of lexical entries (McCrae et al., 2017). The core of the Ontolex-Lemon model is provided in Figure 3. In addition, we also use the lexicography module–Lexicog which provides a conceptual model of language and linguistic objects in lexicography (Bosque-Gil et al., 2017).
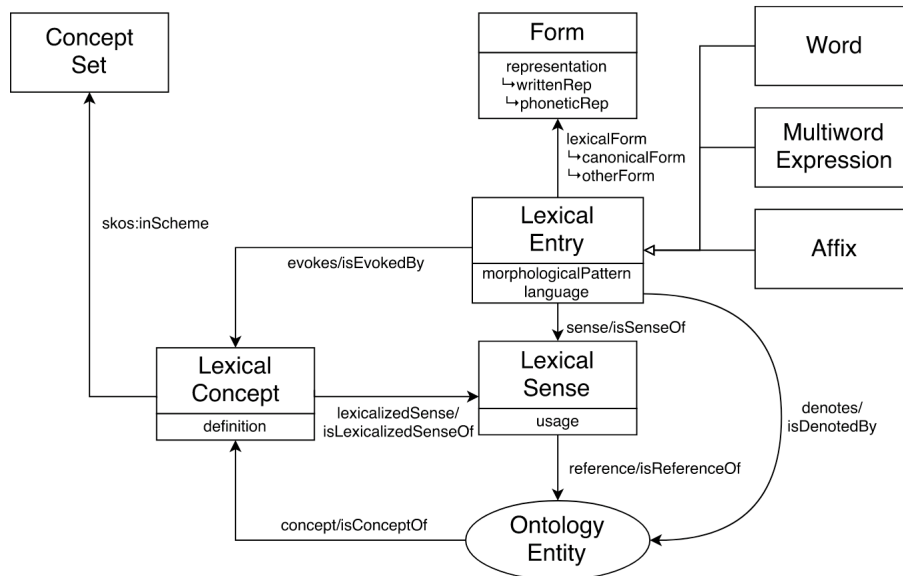

Figure 3: Lemon-OntoLex Core (McCrae et al., 2017)

Figure 4 shows the same entry in Figure 2 where the lemma in the Arabic-based and Latin scripts are provided along with the senses and their translations into Persian. It should be noted that morphosyntactic information such as part-of-speech tags are not provided in the current version of the electronic dictionary. Due to the inconsistency in differentiating glosses and senses in the microstructure of the printed dictionary, we only include senses which are composed of at most two space-separated words. This measure was taken to only include senses rather than glosses in the converted dictionary. Overall, 14,326 entries are extracted from the printed lexicon.

```
1   @prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#> .
2   @prefix vartrans: <http://www.w3.org/ns/lemon/vartrans#> .
3   @prefix lime: <http://www.w3.org/ns/lemon/lime#> .
4   :lexicon a lime:Lexicon;
5       lime:language <www.lexvo.org/page/iso639-3/sdh> ;
6       lime:entry :lex_qirtan .
7
8   :lex_qirtan a ontolex:LexicalEntry, ontolex:Word ;
9       dct:language <www.lexvo.org/page/iso639-3/sdh> ;
10      rdfs:label "qirtan"@sdh-latn ;
11      rdfs:label "قرتان"@sdh-arab .
12      ontolex:sense :qirtan_sense_1, qirtan_sense_2, qirtan_sense_3, qirtan_sense_4,
            qirtan_sense_5 .
13
14  :qirtan_sense_1 rdfs:label "پەرانن"@sdh-arab .
15  :qirtan_sense_2 rdfs:label "چەوەقڕی"@sdh-arab .
16
17  :fa_lex_1 a ontolex:LexicalEntry ;
18      rdfs:label "پراندن"@fa ; ontolex:sense :fa_lex_1_sense .
19  :fa_lex_2 a ontolex:LexicalEntry ;
20      rdfs:label "چشمک زدن"@fa ; ontolex:sense :fa_lex_2_sense .
21
22  :trans_qirtan_sense_1_fa_lex_1 a vartrans:Translation ;
23    vartrans:source :qirtan_sense_1 ; vartrans:target :fa_lex_1 .
24  :trans_qirtan_sense_2_fa_lex_2 a vartrans:Translation ;
25    vartrans:source :qirtan_sense_2 ; vartrans:target :fa_lex_2 .
```

Figure 4: An example entry from our Southern Kurdish dictionary. The original printed entry in the left and the equivalent in RDF Turtle based on the OntoLex-Lemon model

3.3 Alignment with a Sorani Kurdish dictionary

Some of the senses in the southern Kurdish dictionary are provided in Sorani Kurdish. Such cases allow us to align the current resource with other existing ones, particularly (Ahmadi et al., 2019)'s Sorani dictionary in Ontolex-Lemon[5]. The latter provides translations in English for Sorani which can also used as a pivot language to align with other lexical resources. More precisely, we first align the entries in the southern Kurdish dictionary by matching senses that appear in the Sorani Kurdish dictionary. Therefore, the initial headwords can be aligned with the English translations of the Sorani lemmas. The alignment of the Southern Kurdish dictionary with the Sorani one yielded 1,047 cross-dialect links.

---

[5] https://github.com/KurdishBLARK/KurdishLex/blob/master/Sorani.ttl

4. Conclusion and Future Work

In this paper, we present a preliminary study to create an electronic lexicon in southern Kurdish using the Ontolex-Lemon ontology. Our primary goal is to shed light on the current state of the southern varieties of Kurdish. We believe that this resource helps the southern varieties of the Kurdish language, which are under-represented, to be documented. Moreover, it will pave the way for further developments for southern Kurdish, in particular in language technology and natural language processing for tasks such as spelling error detection and correction, part-of-speech tagging and syntactic analysis.

A major limitation of this work is due to the limited coverage of the dictionary and also, lack of glosses, examples, pronunciations and morphosyntactic properties. The current dictionary can be further completed by adding morphosyntactic information, etymological and usage examples. In order to increase inter-operability among Kurdish resources, it is also suggested to align the resource with other lexical and semantic resources such as KurdNet–the Kurdish wordNet Aliabadi et al. (2014) and more dialects of Kurdish.

5. Acknowledgments

6. References

Abdulrahman, R.O., Hassani, H. & Ahmadi, S. (2019). Developing a fine-grained corpus for a less-resourced language: the case of Kurdish. *arXiv preprint arXiv:1909.11467*.

Ahmadi, S. (2020). Building a Corpus for the Zaza–Gorani Language Family. In *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*. pp. 70–78.

Ahmadi, S., Hassani, H. & McCrae, J.P. (2019). Towards electronic lexicography for the Kurdish language. In *Proceedings of the sixth biennial conference on electronic lexicography (eLex)*. eLex 2019.

Ahmadi, S., McCrae, J.P., Nimb, S., Khan, F., Monachini, M., Pedersen, B., Declerck, T., Wissik, T., Bellandi, A., Pisani, I., Troelsgård, T., Olsen, S., Krek, S., Lipp, V., Váradi, T., Simon, L., Gyorffy, A., Tiberius, C., Schoonheim, T., Ben Moshe, Y., Rudich, M., Abu Ahmad, R., Lonke, D., Kovalenko, K., Langemets, M., Kallas, J., Dereza, O., Fransen, T., Cillessen, D., Lindemann, D., Alonso, M., Salgado, A., Luis Sancho, J., Ureña-Ruiz, R.J., Porta Zamorano, J., Simov, K., Osenova, P., Kancheva, Z., Radev, I., Stanković, R., Perdih, A. & Gabrovsek, D. (2020). A Multilingual Evaluation Dataset for Monolingual Word Sense Alignment. In *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, pp. 3232–3242. URL https://www.aclweb.org/anthology/2020.lrec-1.395.

Aliabadi, P., Ahmadi, M.S., Salavati, S. & Esmaili, K.S. (2014). Towards building kurdnet, the kurdish wordnet. In *Proceedings of the Seventh Global Wordnet Conference*. pp. 1–6.

Aliakbari, M., Gheitasi, M. & Anonby, E. (2015). On language distribution in Ilam province. *Iranian studies*, 48(6), pp. 835–850.

Belelli, S. (2019). Towards a Dialectology of Southern Kurdish: Where to begin? In *Current issues in Kurdish linguistics*. University of Bamberg Press, pp. 73–92.

Bosque-Gil, J., Gracia, J. & Montiel-Ponsoda, E. (2017). Towards a Module for Lexicography in OntoLex. In *LDK Workshops*. pp. 74–84.

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

Hassanpour, A. (1992). *Nationalism and language in Kurdistan*. New York: Edwin Mellen Pr.

Izady, M. (1992). *The Kurds: A concise handbook*. London: Taylor & Francis.

Jalilian, A. (2006). *Farhang-ī bāšūr: Tāībat ba nāwčakānī kirmāšān-u īlām-u luřistān (Bashur dictionary: Specific for the regions of Kemanshah, Ilam and Lorestan)*. Tehran: Porseman.

Karimpour, K. (2003). *Khovar Halat: Farhang-e guyeš-e kordi-e kalhori (kermānšāhi), kordi-fārsi (Dictionary of the Kalhori Kurdish (Kermanshahi) dialect, Kurdish-Persian)*. Kermanshah: Sobh-e Roshan.

Kreyenbroek, P. & Chamanara, B. (2013). Literary Gurāni: Koinè or Continuum? *Chez les Kurdes*, pp. 151–169.

McCrae, J.P., Bosque-Gil, J., Gracia, J., Buitelaar, P. & Cimiano, P. (2017). The Ontolex-Lemon model: development and applications. In *Proceedings of eLex 2017 conference*. pp. 19–21.

Nielsen, F. (2020). Lexemes in Wikidata: 2020 status. In *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*. pp. 82–86.

Sharafkandi, A. (1991). *Henbane Borîne Farhang-e Kordi-Farsi (Henbane Borine Kurdish-Persian dictionary)*. Tehran: Soroush Press.

Yarahmadi, J. (2021). Language Shift Among Speakers of Kalhuri Kurdish in Iran. *International Journal of Kurdish Studies*, 7(1), pp. 82–102.