Kurdish and Persian: Dialects or Separate Languages?

Mohammed Khalid Mahmood[1], Ali Qadir Hama Rash[2], Mohammed Amin Qadir Hama Rash[3], Saeed Khalid Mahmood[4], Hüseyin Güler[5]

[1]Komar University of Science and Technology, Kurdistan, Sulaimani, Iraq.
[2]Koya University, Kurdish Studies MSc
[3]Koya University, Arabic Studies MSc.

ABSTRACT: Although it is not quietly supported nowadays, some linguistics and orientalists believed/proposed that due to the large resemblance between the two and somehow mutual ineligibility, Kurdish is a dialect of Persian. The objective of this paper was to answer this hypothesis from a lexical point of view. Lexical Similarity measures the similarity and/or difference between a set of words from any given two languages. Despite the abundance of lexical similarity coefficients between various world languages in the literature, there are no available data on Kurdish and Persian, even though many Kurdish natives learn Persian as a second language. Our results showed (51.5 %) lexical similarity between Persian and Kurdish by using the Leipzig-Jakarta list. This result shows close proximity and common genetic roots between these two languages. However, it is not high enough to claim a dialect relation between the two; like, for example, 85% similarity index between Catalan and Spanish. Hence, this outcome, in a lexical point of view, is against the hypothesis of 'Kurdish is a dialect of Persian'.

KEYWORDS: Kurdish, Persian, Language, Dialect, Lexical Similarity, Indo-Iranian, Indo-European.

INTRODUCTION

Indo-European languages are the native languages of the habitants of south and west Eurasia. With 3.5 billion speakers, it is the largest family of spoken languages in the world, making up 46% of the global population in a broad area spanning from the Indian subcontinent to the Iranian plateau and Europe. It is thought that all Indo-European languages descended from a single common ancestor known as proto-Indo-European (PIE). PIE is assumed to have been spoken between 4500 and 2500 BC, primarily in the Eastern European Caspian region, throughout the Neolithic and Bronze periods (Baldi and Cuzzolin 2018).

Persian and Kurdish language belongs to the western Iranian group of the Indo-Iranian branch of the Indo-European family. Is Persian predominantly spoken and used officially within Iran, Afghanistan, and Tajikistan in three mutually intelligible standard varieties with more than 100 million native speakers. On the other hand, more than 40 million Kurds communicate in Kurdish in four major countries: Turkey, Iraq, Iran and Syria. The main dialects include Northern (also known as Kurmanji), Central (also known as Sorani) and Southern (MacKenzie 1962). Persian and Kurdish are genetically related, belong to the same family branch of languages, and are believed to have evolved from a common proto-language.

The notion that the Iranian languages can be separated into an eastern and a western subgroup since the Middle Iranian periods is one of the most widely recognized conclusions of Iranian dialectology.; to the latter belong Persian, Kurdish, Balōči, and all other Iranian languages and dialects that are spoken in present-day Iran (e.g., Gilaki, Tāleši, etc.) (KALA 2020).

Unfortunately, studies on Kurdish language and its history are few and not complete. In addition, many bias has been introduced into this domain due to politicization of every notion connected to the Kurds both pros and contras (Asatrian 2009). In 1961, MacKenzie made the pivotal step in the study of Kurdish history and western Iranian dialectology in general. In his research, he created a complex dialectological model that claimed Kurdish was quite close to Persian after a select number of isoglosses (MacKenzie 1961). MacKenzie also published his dissertation on the Kurdish dialects (and their sub-grouping) that year, and it has since become a widely used resource.(MacKenzie 1962).

Gernot Windfuhr validated the close relationship between Persian and Kurdish, and also Balōči (G.L. Windfuhr 1975); Pierre Lecoq analyzed several phonological and other grammatical aspects, which was appreciated (Lecoq 1989). He also noted that Kurdish and the dialects of Central Iran were more closely related, although he did not provide a detailed account of how this genealogical relationship differed from MacKenzie's.

It is uncertain if it will ever be able to create a thorough historical grammar of western Iranian languages given the scarcity of historical material and the predominance of areal, loan, and substratum components. According to rigorous definitions, this kind of grammar is a requirement for a complete account of Kurdish history. Another step toward constructing such a historical grammar

Kurdish and Persian: Dialects or Separate Languages?

is the following attempt to summarize the most significant and well-known facts about the development of Kurdish (G. Windfuhr 2013).

Although it is not quietly supported nowadays, some linguistics and orientalists believed/proposed that due to the large resemblance between the two and somehow mutual ineligibility, Kurdish is a dialect of Persian (Edmonds 2013). Even Mackenzie says "my first task then should be to define Kurdish, by establishing the features which distinguish it from other Iranian dialects. Unfortunately, I have to admit at the outset that my findings are largely negative, for almost every feature of Kurdish has its counterpart in at least one other Iranian dialect." (MacKenzie 1961) After this statement, he gives only four morphological and phonetic developments that are distinct to Kurdish.

Here, we believe it is useful to clarify what we mean by the terms of 'language' and 'dialect'. Language is an abstract system of signs and meanings that is guided by grammatical rules. Contrarily, dialects are social or geographical variations of a language that are identified by their differences in vocabulary, phonetics, and/or grammar (Haugen 1966).

Lexical Similarity

Similar words (cognates) may appear among different languages mainly due to etymological roots from a common parent language and simple borrowing between languages (Blasi et al. 2016). Lexical Similarity (LS) measures similarity and/or difference between a set of words from any given two languages. A lexical similarity of 1 (or 100%) means a total resemblance between vocabularies, whereas 0 means there are no common words; hence, ranging from mutual intelligibility to lack of relation. A degree of more than 85% strongly suggests a dialect relationship between the two (Do et al. 2009).

There are several ways to estimate LS between two certain languages. Historically, various lists have been proposed by experts like (Swadesh 1955), (Bender 1975) and (Haspelmath 2008). Lexical cognates are used to create the lists used to calculate lexical similarity. Various researchers have different ideas about what cognates are and how to define them. Some think that two pairs of words must share orthographic, phonetic, and semantic similarities in order to be considered cognates, while others look for etymological evidence from a parent language (Otwinowska 2015).

The key feature of the lists suggested for lexical similarity is that they comprise terms that are least likely to have been acquired from other languages and most resistant to change through time. This is why the majority of these lists consist of names of bodily parts, animals, natural elements of the environment, and verbs used in everyday life. Since these names were important for the earliest speakers of any language from the beginning and they had plenty of time to establish deeper cultural roots, they are less susceptible to change (Holman et al. 2011).

There are plenty of papers in the literature on the lexical similarity between languages, especially concerning English and Roman languages. Table 1 shows the lexical comparison between several Indo-European languages. In addition, previously we have shown that the degree of lexical similarity between Kurdish and English is 10%, indicating a remote genetic root between the two (Mahmood). The objective of this paper is to test the hypothesis of 'Kurdish is a dialect of Persian' by measuring the degree of lexical similarity between these two languages/dialects; a task which is never done before.

Table 1: Lexical similarity between selected Indo-European languages (Source: *Ethnologue.com*)

	Catalan	English	French	German	Italian	Portuguese	Romanian	Romansh	Russian	Sardinian	Spanish
Catalan	1	-	0.85	-	0.87	0.85	0.73	0.76	-	0.75	0.85
English	-	1	0.27	0.60	-	-	-	-	0.24	-	-
French	0.85	0.27	1	0.29	0.89	0.75	0.75	0.78	-	0.80	0.75
German	-	0.60	0.29	1	-	-	-	-	-	-	-
Italian	0.87	-	0.89	-	1	-	0.77	0.78	-	0.85	0.82
Portuguese	0.85	-	0.75	-	-	1	0.72	0.74	-	-	0.89
Romanian	0.73	-	0.75	-	0.77	0.72	1	0.72	-	0.74	0.71
Romansh	0.76	-	0.78	-	0.78	0.74	0.72	1	-	0.74	0.74
Russian	-	0.24	-	-	-	-	-	-	1	-	-
Sardinian	0.75	-	0.80	-	0.85	-	0.74	0.74	-	1	0.76
Spanish	0.85	-	0.75	-	0.82	0.89	0.71	0.74	-	0.76	1

METHOD

Several methods have been used to compute lexical similarity between languages by (Holmes and Ramos 1993), (Kessler 1995) and (List 2012). The main idea is presented as the following: "Cognate identification is usually based on a similarity or distance score

calculated from the number of matches and mismatches in the alignment" (List 2012). The scoring system of this research is mainly based on the Levenshtein algorithm by the (Kessler 1995).

In this survey, the 100 words of Leipzig-Jakarta list (Haspelmath 2008) by the Max Planck Digital Library is used as references (Haspelmath 2008).

For Persian, the Iranian Farsi is taken into consideration. Beside the formal Arabic-derived Persian alphabet, a Latinized version is provided according to the *journal of Persianate Studies* (Arjomand 2015). Table 2 shows the summary of this Latinization.

Table 2: Latinized Persian alphabet

Consonants		Vowels	
ء	'	صَ ◌ٰ	a
ب	b	◌ِ	e
پ	p	◌ُ	o
ت	t	ا-/آ-	ā
ث	s (or th)	ی	i
ج	j	و	u
چ	ch	◌ِی	ey (or ay)
ح	h	◌َو	ow
خ	kh		
د	d		
ذ	z (or dh)		
ر	r		
ز	z		
ژ	zh		
س	s		
ش	sh		
ص	s		
ض	z		
ط	t		
ظ	z		
ع	'		
غ	gh		
ف	f		
ق	q		
ک	k		
گ	g		
ل	l		
م	m		
ن	n		
و	v (w)		
ه	h		
ی	y		

For Kurdish, two main dialects of North and Central Kurdish which are spoken by the majority of the native Kurds are used in the comparison. The North dialect (also known as Kurmanji) already uses an adapted Latin alphabet; however, the Central dialect (also known as Sorani) uses an adapted Arabic alphabet. For the purpose of easier visual comparison, words of the Sorani dialect are also converted into the Latin alphabet. Table 3 shows information on the adapted Latin alphabet of Kurdish language.

Table 3: Main characteristics of the adapted Kurdish alphabet

Kurdish	Correspondent in English
A	Long a like in art
b, d, f, g, h, k, l, m, n, o, p, r, s, t, u, v, w, y, z	Same as English
C	like /dj/ in June
Ç	/ch/ like in chain
E	short a like /a/ in any
Ê	/e/ like in leg
i	like the short unwritten vowel between s and t in star
î	/i/ like in list
j	/zh/ like in the French j
q	[q] like Arabic (ق)
ş	/sh/ like in shame
x	[kh] like Arabic (خ)

After enlisting the words, the sound and the orthography are compared. Five categories of scorings are applied. (0) denotes no resemblance between the two pairs (Ex: *bini/lut*). (0.25) is the score for slight similarity between the words, accounting for one or two letters of the word according to the word size (Ex: *ātash/agir*). (0.5) is the score for a similarity that exist in half of the word or the sound (Ex: *āb/aw*). (0.75) is for an apparent resemblance between the sound of the pairs with a slight difference (Ex: *zabān/ziman*). Score of (1) denotes a total similarity between the words (Ex: *tu/tu*).

In certain instances, the similarity between one of the dialects and Persian was more obvious and striking when compared to its counterpart dialect. In this instance, the one with the greater degree of similitude is taken into parenthesis and used as a reference in the comparison.

Kurdish and Persian: Dialects or Separate Languages?

RESULTS

Majority of the words (80%) revealed some kind of similarity. Table 4 shows the complete Leipzig-Jakarta list and the scores and Table 5 shows the summary of the results.

Regarding the scoring calculation technique, some of the previous researches and the data from Ethnologue.com have calculated the number of words with positive lexical similarity (regardless their degree) divided by the whole number of words studied in the survey. Hence, according to this approach, the degree of lexical similarity between Kurdish and Persian is (80%). However, in our previous study we proposed a more accurate methodology which is based on the exact similarity score of each word (Mahmood). Eventually, a total degree of (51.5%) of similarity which is the addition of all the scores is observed between Kurdish and Persian using the Leipzig-Jakarta list.

Table 4: Results of the Leipzig-Jakarta list

No	English	Persian	Kurdish	Score
1.	Fire	آتش (ātash)	agir	0.25
2.	Nose	بینی (bini)	poz, lut	0
3.	To go	رفتن (raftan)	çun, (roîştin)	0.5
4.	Water	اب (āb)	av, aw	0.5
5.	Mouth	دهان (dahān)	dev, dem	0.25
6.	Tongue	زبان (zabān)	ziman	0.75
7.	Blood	خون (khun)	xwîn/xwên	0.75
8.	You (singular)	تو (tu)	(tu), to	1
9.	Root	ریشه (risha)	reh, reg	0.25
10.	To come	آمدن (āmadan)	hatin	0.25
11	Bone	استخوان (istekhān)	hestî, êsk, (êsqan)	0.5
12	Breast	پستان (pastān) سینه (sina)	(pêşîr), memk	0.25
13	Rain	باران (bārān)	baran	1
14	I/me	من (man)	ez, (min)	1
15	Name	نام (nām)	nav, naw	0.75
16	Louse	شپش (shish)	spî, espê	0
17	Wing	بال (bāl)	bask,(bal)	1
18	Flesh/meat	گوشت (gusht)	goşt	0.75
19	Hand	دست (dast)	dest	1
20	Fly	مگس (magas)	mêş	0.25
21	Night	شب (shab)	şev, şew	0.75
22	Ear	گوش (gush)	guh, gwê	0.75
23	Neck	گردن (gardan)	hustu, mil	0
24	Far	دور (dur)	dur	1
25	To do	کردن (kardan)	Kirin, (kirdin)	1
26	House	خانه (khāna)	xanî, xanu	0.75
27	Stone	سنگ (sang)	Kevr, berd	0
28	Bitter	تلخ (talkh)	tal	0.5
29	To say	گفتن (guftan)	(gotin), witin	0.5
30	Tooth	دندان (dendān)	diran, dan	0.75
31	Hair	مو (mu)	por, qij, (mu)	1
32	Big	بزرگ (bozorg)	mezin/gewre	0
33	One	یک (yak)	yek	1
34	Who?	کی (ki)	(kî), kê	1
35	He/she/it	او (u)	ew	0.75
36	To beat/hit	زدن (zadan)	lêdan	0.5
37	Leg/foot	پا (pā)	ling, qaç, (pê)	0.5
38	Horn	شاخ (shākh)	qiloch, (şax)	1
39	This	این (in)	ev, em	0.25
40	Fish	ماهی (māhi)	masî	0.75
41	Yesterday	دیروز (diruz)	do, dwênê	0.25
42	To drink	نوشیدن (nushidan)	vexwarin, xwardinewe	0.25

43	Black	سیاه (siyāh)	reş	0
44	Navel	ناف (nāf)	navik, nawik	0.5
45	To stand	ایستادن (istādan)	rawestîn, westan	0.75
46	To bite	گاز زدن (gāz zadan)	gezkirin, (gazgirtin)	0.5
47	Back	پشت (pusht)	(paş), dwa	0.5
48	Wind	باد (bād)	ba	0.75
49	Smoke	دود (dud)	dukel	0.5
50	What?	چی (chi)	çî	1
51	Child	کودک (kudak) بچه (bacha)	zarok, minal	0
52	Egg	تخم (tokhm)	hêk, hêlke	0
53	To give	دادن (dādan)	(dan), pêdan	0.5
54	New	نو (nu)	(nu), nwê	1
55	To burn (intransitive)	سوختن (sukhtan)	sotin, sutān	0.75
56	Not	نه (na)	Na, ne	1
57	Good	خوب (khub)	Baş	0
58	to know	دانستن (dānestan)	zanîn	0.25
59	Knee	زانو (zānu)	çok, ejno	0.25
60	Sand	شن (shan)	xîz, lim	0
61	To laugh	خندیدن (khandidan)	kenîn, pêkenîn	0.25
62	To hear	شنیدن (shanadan)	bihîstin, bîstin	0.25
63	Soil	خاک (khāk)	ax, (xak), zewî	1
64	Leaf	برگ (barg)	(belg), pel, gela	0.75
65	Red	سرخ (sorkh)	Sor, sur	0.75
66	Liver	جگر (jegar)	Jeger, Jiger	1
67	To hide	نهفتن (nehaftan)	(veşartin), şardinewe	0.25
68	Skin	پوست (pust)	çerm, (pêst)	0.75
69	To suck	مکیدن (makidan)	mijîn	0.25
70	To carry	حمل کردن (haml kardan) بردن (burdan)	Helgirtin	0
71	Ant	مورچه (murcha)	mîrok, mêrule	0.5
72	Heavy	سنگین (sangin)	giran, qurs	0
73	Wood	چوب (chub)	texte	0
74	To take	گرفتن (gereftan) بردن (burdan)	girtin, birdin	0.75
75	Old	کهنه (kohna)	kevin, kon	0.75
76	To eat	خوردن (khurdan)	xwarin, xwardin	0.75
77	Thigh	ران (rān)	ran	1
78	Thick	کلفت (kolaft)	estur	0
79	Long	دراز (derāz)	dirêj	0.5
80	To blow	دمیدن (damidan)	pifkirin, futêkirdin	0
81	To run	دویدن (davidan)	(revîn), rakirdin	0.25
82	To fall	افتادن (uftādan)	ketin, kewtin	0
83	Eye	چشم (chashm)	çav, çaw	0.5
84	Ash	خاکستر (khākestar)	xwalî, xolemêş	0.25
85	Tail	دم (dam)	terrî, poçik, kilk	0
86	Dog	سگ (sag)	seh, (seg)	1
87	To cry/weep	گریه کردن (gerya kardan)	girîn, giryan	0.5
88	To tie	بستن (bastan)	girêdan, (bestin)	1

89	To see	دیدن (*didan*)	(*dîtin*), *bînîn*	0.75
90	Sweet	شیرین (*shirin*)	*Şîrîn*	1
91	Rope	ریسمان (*rismān*) رسن (*rasan*)	*werîs, gurîs*	0.5
92	Shadow	سایه (*sāya*)	*sî, sêber*	0.25
93	Bird	پرنده (*paranda*)	*balinde*	0.75
94	Salt	نمک (*nemak*)	*xwê*	0
95	Small	کوچک (*kuchek*)	*biçuk*	0.5
96	Wide	پهن (*pehan*)	*pan*	0.75
97	Star	ستاره (*setāra*)	*sitêrk, (estêre)*	0.75
98	In	در (*dar*) از (*az*)	*li, le*	0
99	Hard	سنگین (*sangin*) سخت (*sekht*)	*req*	0
100	to chew	جویدن (*juidan*)	*cutin, (cwîn)*	0.75

Table 5: Summary results

Score	Number of words	Value
0	20	0
0.25	18	4.5
0.5	18	9
0.75	24	18
1	20	20
Total	**100**	**51.5**

CONCLUSION AND PRACTICAL IMPLICATIONS

In this study, for the first time in the literature, Persian and Kurdish languages are studied in relation to the degree of lexical similarity by using Leipzig-Jakarta list. Result of this paper (51.5%) show close proximity and common genetic roots between these two languages. However, it is not high enough to generate mutual intelligibility between the speakers of the two; like, for example, 85% similarity index between Catalan and Spanish. Hence, this result, in a lexical point of view, is against the hypothesis of 'Kurdish is a dialect of Persian'. Also, it should be noted that the result of (51.5%) should be seen as a maximum score because we have chosen the synonym with the highest level of resemblance and neglected the other one(s) in the comparison.

Practically speaking, being aware of these similarities helps Persian and Kurdish second-language students and teachers better understand the connections between the two languages, and feel motivated to keep learning and teaching.

REFERENCES

1) Arjomand, Saïd Amir. 2015. "A Decade of Persianate Studies." *Journal of Persianate Studies* 8 (2): 309-333.
2) Asatrian, Garnik. 2009. "Prolegomena to the Study of the Kurds." *Iran and the Caucasus* 13 (1): 1-57.
3) Baldi, Philip, and Pierluigi Cuzzolin. 2018. "Indo-european languages." In *The World's Major Languages*, 23-50. Routledge.
4) Bender, M Lionel. 1975. "Toward a Lexicostatistic Classification of Ethiopian Languages." *Bynon and Bynon, eds*: 377-88.
5) Blasi, Damián E, Søren Wichmann, Harald Hammarström, Peter F Stadler, and Morten H Christiansen. 2016. "Sound–meaning association biases evidenced across thousands of languages." *Proceedings of the National Academy of Sciences* 113 (39): 10818-10823.
6) Do, Quang, Dan Roth, Mark Sammons, Yuancheng Tu, and V Vydiswaran. 2009. "Robust, light-weight approaches to compute lexical similarity." *Computer Science Research and Technical Reports, University of Illinois* 9.
7) Edmonds, Alexander Johannes. 2013. "The Dialects of Kurdish." *Ruprecht-Karls-Universität Heidelberg*.
8) Haspelmath, Martin. 2008. "Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability." *Empirical Approaches to Language Typology* 35: 43.
9) Haugen, Einar. 1966. "Dialect, language, nation 1." *American anthropologist* 68 (4): 922-935.

Kurdish and Persian: Dialects or Separate Languages?

10) Holman, Eric W, Cecil H Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, and Pamela Brown. 2011. "Automated dating of the world's language families based on lexical similarity." *Current Anthropology* 52 (6): 841-875.

11) Holmes, John, and Rosinda G Ramos. 1993. "False friends and reckless guessers: Observing cognate recognition strategies." *Second language reading and vocabulary learning*: 86-108.

12) KALA, JULFA I SAFAVID PERIOD–KAFIR. 2020. "Encyclopaedia Iranica."

13) Kessler, Brett. 1995. "Computational dialectology in irish gaelic." *arXiv preprint cmp-lg/9503002*.

14) Lecoq, Pierre. 1989. "Les dialectes du centre de l'Iran." *Compendium linguarum iranicarum*: 311-326.

15) List, Johann-Mattis. 2012. "LexStat: Automatic detection of cognates in multilingual wordlists." Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH.

16) MacKenzie, David Neil. 1961. "THE OKIGINS OF KURDISH." *Transactions of the Philological Society* 60 (1): 68-86.

17) ---. 1962. *Kurdish Dialect, Studies 1*. Vol. 2. Oxford University Press.

18) Mahmood, Mohammed Khalid. "Degree of Lexical Similarity between English and Kurdish."

19) Otwinowska, Agnieszka. 2015. *Cognate vocabulary in language acquisition and use*. Multilingual Matters.

20) Swadesh, Morris. 1955. "Towards greater accuracy in lexicostatistic dating." *International journal of American linguistics* 21 (2): 121-137.

21) Windfuhr, Gernot. 2013. "Introduction to the iranian languages." In *The Iranian Languages*, 77-80. Routledge.

22) Windfuhr, Gernot L. 1975. *Isoglosses: a sketch on Persians and Parthians, Kurds and Medes*. na.