



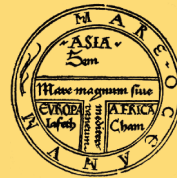
ISTITUTO INTERNAZIONALE DI CULTURA KURDO  
ISMEO – ASSOCIAZIONE INTERNAZIONALE DI STUDI  
SUL MEDITERRANEO E L'ORIENTE

Sara Beelli

# For a Kurdish-Italian Dictionary



Istituto  
kurdo



Roma 2020





## 1. Introduction

In the context of Kurdish lexicography—whose history is recent but prolific in almost all major European languages—the publication of the first modern Kurdish-Italian bilingual dictionary is perceived as an urgent desideratum. Due to the current lack of teaching resources for Kurdish in Italian, this would represent an important first reference for Italian speakers willing to study the language, as well as for Kurdish speakers engaged in acquiring Italian as a second language.

The completion of this work has been hoped for and outlined several times, at least since the manuscripts of Coletti's (1985) *Dizionario curdo (sorani, mukri, kurmanji) italiano, pratico, etimologico, comparativo*, but a Kurdish-Italian dictionary would currently need to be planned within a more systematic and project-oriented framework. Indeed, the peculiar sociolinguistic and dialectological status of Kurdish makes it a special case in lexicography and calls for additional caution compared to other language candidates for a lexical collection aimed at practical use.

Indeed, 'Kurdish' can be used as language label to denote a multitude of referents. It is commonly applied to a very diverse group of varieties that contemporary dialectology divides into three main subgroups, listed below by decreasing number of speakers: Northern Kurdish (Kurmanji, Bahdini, etc.); Central Kurdish (Sorani, Mukri, Sina'i, etc.); Southern Kurdish (Kalhori, Kermanshahi, Ilami, etc.). The differences between these macro-varieties can be significant enough to complicate verbal interactions between speakers of geographically distant locations or peripheral regions of the continuum, which according to recent studies should rather be described as an 'area of linguistic convergence.'

Although variation concerns morphosyntax in the first place—e.g., the presence or absence of grammatical gender, case and enclitic pronouns; or divergences in the conjugation of past transitive verbs—lexicon is not immune to variation, both within and across dialect clusters. The use of different graphic systems is also current, with a Latin-based

<sup>1</sup> An 'area of linguistic convergence' or 'linguistic area' (German *Sprachbund*) can be defined as "the solidarity established in a geographical area and a specific historical phase between genetically unrelated languages leading to convergence phenomena, such as between various Balkan, Finno-Ugric, Slavic, Romance languages, etc." (Nuovo De Mauro online <https://dizionario.internazionale.it/parola/lega-linguistica>). This happens as a result of constant and prolonged mutual influence, possible also thanks to common belonging to the same cultural area.

script used for Kurmanji<sup>2</sup> and an Arabic-based script used for Sorani<sup>3</sup> besides other systems (“Ottoman” Arabic, Cyrillic, Armenian) historically employed for northern varieties on a regional basis.

Therefore, a preliminary definition of the target variety of our collection appears as an essential requirement. Especially because the aim is producing a lexical repertoire suitable for practical use by non-specialists, that should reflect a *tangible* linguistic reality beyond all areal diversity and an imperfect and uneven standardization, as observed for both majority varieties of Kurdish (Kurmanji and Sorani).

To date, Sorani seems the most suitable choice according to the needs of Italian-speaking and Kurdish-speaking readers in Italy. It would also seem to offer better chances for the elaboration of a product that meets the criteria of modern lexicography, requiring the availability of sufficient reference samples and language corpora diversified by register.

Indeed, thanks to the efforts made in recent decades by academic and political institutions of the Kurdistan Regional Government, Kurdish cultural institutions around the world, scholars in Kurdish Studies and—not less importantly—native speakers in the motherland and the diaspora, Sorani has made huge progress in lexical and grammatical systematization.<sup>4</sup> So much so that it shows ever-increasing applicability in the fields of administration, academia, technology, media communication, and literature. Sorani also boasts a broad diffusion in spoken and written use, often surpassing the territorial borders of the Republic of Iraq, where since 2005—for the first time in modern history—Kurdish holds the status of official language alongside Arabic.<sup>5</sup>

## 2. Kurdish Sorani Bilingual Lexicography in European Languages

Sorani has been codified as a literary language only in the 19<sup>th</sup> century, based on the variety of Sulaymaniya (Kurd. سلیمانی *Silêmani*). This means that this variety has a much later written tradition compared to that of Kurmanji Kurdish, for which the appearance of relevant literary oeuvres dates back to the 16<sup>th</sup> and 17<sup>th</sup> centuries.<sup>6</sup> These early works have undoubtedly paved the way for later Kurdish literary production—primarily conveyed through magazines such as *Hawar* (The Cry)—as well as for the development of

<sup>2</sup> The extended Latin script of Kurmanji, known as the “Hawar alphabet” or “Bedir Khan alphabet,” was developed based on the Turkish model around the 1830s. It was later popularized by the *Hawar* periodical founded by the Bedir Khan brothers. For an overview of the processes leading to the elaboration of a Kurdish written norm, cf. A. V. Rossi, 2007, “La formazione delle lingue iraniche tra stratificazione multiculturale e pianificazione linguistica,” *Atti del XXX Convegno annuale SIG “Standard e non standard tra scelta e norma” (Bergamo 20-22 ottobre 2005)*, Roma: Il Calamo, pp. 91-97.

<sup>3</sup> A modified Arabic script was developed and introduced for Sorani by Tawfiq Wahby since the 1920s. For a summary of the historical phases leading to the standardization of the Arabic-based script, cf. A. Hassanpour, 1992, *Nationalism and Language in Kurdistan 1918-1985*, San Francisco: Mellen Research University Press, pp. 357-369.

<sup>4</sup> For a comprehensive history of Kurdish standardization processes, see A. Hassanpour, *op. cit.*

<sup>5</sup> The Constitution of the Republic of Iraq, Article 4, states that “the Arabic language and the Kurdish language are the two official languages of Iraq. The right of Iraqis to educate their children in their mother tongue, such as Turkmen, Assyrian, and Armenian shall be guaranteed in government educational institutions in accordance with educational guidelines, or in any other language in private educational institutions.”

the first modern Kurdish grammars and lexical repertoires, usually compiled by the same intellectuals engaged in the publication of Kurmanji periodicals.

Among these intellectuals, stand out the names of the brothers Celadet and Kamuran Ali Bedir Khan. We owe to the latter the ambitious project for a Kurdish-French dictionary imagined as a comprehensive repertoire of the vocabulary of all languages of Kurdistan, including Zazaki/Dimli and Gorani/Hawrami, that was to replace A. Jaba & F. Justi's *Dictionnaire kurde-français* (1879, St. Petersburg: Académie impériale des sciences), a contradictory product of nineteenth-century philology. This project, later supported by the Institut kurde de Paris, has been recently brought to completion under the direction of Kendal Nezan with the publication of his *Dictionnaire kurde-français/Ferhenga kurdi-fransîzî* (2017, Paris: Riveneuve Editions, 85,000 entries). Although this dictionary is essentially based on Northern Kurdish,<sup>7</sup> it includes many areally-defined lexical variants among which a good number of Sorani entries marked as SOR.

The success of this large-scale project—almost unique in terms of size in the context of modern bilingual Kurdish dictionaries—owes much to Soviet contributions to Kurmanji lexicography. Among these should be mentioned Ç. X. Bakaev's *Kurdsko-russkij slovar'* (1957, Moscow: Gosudarstvennoe izdatel'stvo, 14,000 entries); I. O. Farizov's *Russko-kurdskij slovar'* (1957, Moscow: Russkij jazyk, 30,000 entries) and K. K. Kurdoev's *Kurdsko-russkij slovar'/Ferhenga Kurdi-Rusî* (1960, Moscow, Gosudarstvennoe izdatel'stvo, 34,000 entries).

As we have seen, the first steps of modern Kurdish lexicography were taken mostly in the context of northern varieties. Even so, thanks to the activity of reformers such as Tawfiq Wahby, Central varieties of Iraqi Kurdistan—under British mandate since the 1920s—could equally develop a grammatical and orthographic norm. The collaboration of the mentioned Kurdish intellectual with the high-ranking Political officer C. J. Edmonds eventually led to the ground-breaking publication of *A Kurdish-English Dictionary* (T. Wahby & C. J. Edmonds, 1966, Oxford: Clarendon Press) bound to remain for a long time the main bilingual reference dictionary for Central Kurdish.<sup>8</sup> The advancement of Sorani lexical studies allowed the Russian school to narrow the gap with Central varieties, with the publication of K. K. Kurdoev & Z. A. Yusupova's *Kurdsko-russkij slovar' (sorani)* (1983, Moscow: Russkij jazyk, 25,000 entries).

<sup>6</sup> We refer in particular to the poetic works of authors such as Eliyê Herîrî (Ali Hariri), Mela Ehmed Cizîrî (Mulla Ahmad Jaziri) and Feqiyê Teyran (cf. A. Hassanpour, op. cit., p. 72ff.). This period of incubation was followed by the publication of Ehmedê Xanî (Ahmadi Khani)'s epic poem *Mem û Zîn* whose appearance in 1685 marks, in common perception, the beginning of the Kurmanji cultural and literary renaissance.

<sup>7</sup> For a bibliography of dictionaries and textual sources variously involved in the compilation of this work, see pp. XLIII-XLVI of the introduction to the dictionary. Bibliographic information on modern Kurmanji bilingual dictionaries not overtly mentioned here can also be found at pp. X-XV of the introduction to M. Chyet's now classic *Kurdish-English Dictionary Ferhenga Kurmancî-Inglîzî* (2003, New Haven/ London: Yale University Press). See also the very recent publication of a revised and expanded re-edition of the work, now integrating the Arabic-based Kurdish script, under the title *Ferhenga Birûskî – Kurmanji-English Dictionary* (M. Chyet, 2020, 3 vols., London: Transnational Press).

<sup>8</sup> The collection is representative of a variety of language described as the “standard language of belles lettres, journalism, official and private correspondence and formal speech as it has been developed, on the basis of Southern-Kirmanji dialect of Suleimani in Iraq since 1918.”

It is also worth mentioning the particular case of *Saladin's English-Kurdish Dictionary/Qamûsa Selahêdîn îngilîzî-kurdî* (S. Saadalla 1998, Baghdad: Dar al-Hurriya; 2<sup>nd</sup> ed. 2000, Istanbul: Avesta; Paris: Institut kurde de Paris), originally published with headwords in Arabic characters and later reprinted in Latin script. About the dictionary, Chyet (2003: XV) writes: "One finds in it many obscure and superfluous English words, some of which are quite amusing [...]. Often the Sorani equivalent is given, rather than the expected Kurmanji word [...]. The dictionary is largely limited to Iraqi Kurdish usage."

More recently, Sorani bilingual lexicography has witnessed the appearance of other works of great relevance, among which S. Qazzaz's *The Sharezoor Kurdish-English Dictionary* (2000, Erbil: Aras Press) and H. Hakem's *Dictionnaire kurde-français (sorani)* (2012, Dictionnaires INALCO, Paris: L'Asiathèque, 22000 entries). The latter follows by many years the publication of the *Dictionnaire français-kurde* (H. Hakem, G. Gautier, 1993, Paris: Klincksieck).

In addition to these, a good number of lexicons and glossaries can be found in the appendix to teaching and descriptive grammar books, among which J. Blau's *Methode de kurde sorani* (2000, Paris/Montréal: L'Harmattan) and W. M. Thackston's *Sorani Kurdish: A Reference Grammar with Selected Readings* (2006, Harvard University).

Some valuable lexicographic tools, generally serving the purpose of second language acquisition, have also appeared in Sweden and Germany among the main target countries of the Kurdish diaspora in Europe. The most relevant titles are the *Kurdisk-svensk ordbok (Sorani)* (M. K. Rashid, 1994, Lund: Studentlitteratur, 10,000 entries); the *Svensk-kurdisk ordbok* (M. K. Rashid, 2<sup>nd</sup> ed., 2007, Stockholm: Alfabet Maxima, 35,000 entries); the *Kurdisch-Deutsches Wörterbuch (Sorani)* (F. F. Omar, 2005, Berlin: Institut für kurdische Studien, 60,000 entries); and the *Deutsch-Kurdisches Wörterbuch (Sorani)* (F. F. Omar, 2016, Berlin: Institut für kurdische Studien, 100,000 entries).

Obviously, the appearance of most bilingual dictionaries mentioned above rests on an extensive Kurdish-Kurdish monolingual lexicography, as well as on Kurdish-Turkish, Kurdish-Arabic and Kurdish-Persian bilingual dictionaries<sup>9</sup> not illustrated in detail here. It is nonetheless worth mentioning the *Ferheng-î xat* (Kurdish-Kurdish, 3 vols., Sheik M. Khal, 1960-1976, Sulaymaniya: Kameran Press) among the primary sources of virtually all modern dictionaries of Sorani Kurdish.

<sup>9</sup> The last two categories are particularly relevant for Sorani, as is natural according to geographic distribution. They include many glossaries for specialized vocabularies (technological, scientific, medical, chemical, agricultural, administrative, etc.). A few representative titles are mentioned in the introductions to S. Qazzaz (2000, pp. 9-10) and H. Hakem (2012, pp. 9-10).

### 3. Project for a Bilingual (Sorani) Kurdish – Italian Dictionary

#### 3.1 Goals

As we have seen, Sorani Kurdish nowadays boasts a very lively narrative, journalistic, academic and literary production, besides being used as a spoken language by no less than 30% of Kurdish speakers in the world. The current situation would call for a constant revision and enlargement of the lexical collections available for this Kurdish variety. The historical tradition of Kurdish studies in Italy,<sup>10</sup> although long interrupted and perhaps never really popularized outside the country, fully justifies the contribution of the Italian language to this common goal.

The aim is to create a medium-sized bilingual lexicographic tool (approximately 25,000-30,000 entries) suitable for practical use by a broad audience, mainly of Italian speakers. The dictionary should reflect as accurately as possible the current lexical repertoire of Sorani Kurdish, understood as the official language of the Republic of Iraq, the administrative language of the Kurdistan Regional Government and the majority language spoken in the territories belonging to the autonomous Kurdish region of Iraq.

The corpus will be collected based on the examination and comparison of monolingual vocabularies and recent bilingual Sorani dictionaries. Given that the balanced corpora and electronic textual banks available for Sorani—unlike those existing for languages with longer written traditions—are both recent and largely untested, the quality of the final work will depend on the ability to make a reasonable use of the lexicographic and technological resources at hand.

Without neglecting the constant verification of data, it will be necessary to refer also to the extensive Kurmanji Kurdish lexicography, as it often contains relevant information regarding forms and usages attributable (also) to Sorani.<sup>11</sup> The work will require a collection of textual and phraseological material from different linguistic registers and domains through printed and online publications, in order to compare it with the evidence of spoken language.

Depending on the project's budget, a working group with diversified skills and mother-tongue collaborators—in any case no less than three people besides the Project Director—will be necessary to identify and organize materials, process data with IT tools, and prepare the final product for the press.

<sup>10</sup> Note that the first description of Kurdish in a European language was compiled in Italian by Maurizio Garzoni (1727, *Grammatica e vocabolario della lingua kurda*, reprinted 1787, Roma: Sacra Congregazione di Propaganda Fide).

<sup>11</sup> As relevantly stated by Chyet (2003: VIII): “While it is now stylish to pepper one’s Kurmanji with Sorani words, this was not the case a decade ago. I have often been in a quandary as to how much of this vocabulary to include.”

### 3.2. Challenges

For all the reasons outlined above, Sorani represents a veritable challenge for modern lexicography. The main critical aspects posed by this project are listed below. For each of them, we will propose and discuss possible solutions.

- 1) Selection of lemmatized words and balance between descriptiveness and normativity
- 2) Orthography and transcriptions
- 3) Structure, compilation and ordering of headwords
- 4) Use of IT tools and lexicographic software.

### 3.3. Selection of Lemmatized Words and Balance between Descriptiveness and Normativity

As we have seen, Sorani has achieved a written codification only in the course of the last century and has become the official language of a geopolitically defined territory no more than twenty years ago. For this reason, written corpora are still overall limited, but at the same time increasingly expanding in virtually all fields, with obvious consequences concerning the creation of neologisms, borrowing processes, the elaboration of technical vocabularies, as well as the presence of nonce words, idiosyncratic forms, orthographic variants, etc.

This background makes a precise definition of the means and purposes of our lexical selection mandatory, seeking a balance between the description of real use and the choice between orthographic and synonymic variants. The latter question concerns, in particular, the contrast between loans and neologisms (cf. the items *ماتەر یالیست* *materyalîst*; *ماتەر یالیزم* *materyalîzm* in §4) or between diverging regional/dialectal forms (cf. *ماتاو* *mataw* vs. *مانگەشەو* *mangêşew* in §4).

We may in principle accept all the items contained in recently published Sorani dictionaries, and later proceed to a comparison, selection and eventual exclusion or integration of materials. The purpose of this task should be, on the one hand, the elimination of possible “ghost words” or obsolete forms borrowed from older lexicography;<sup>12</sup> on the other, the meticulous evaluation of technical terms recently come into use, by omitting forms that are too marginal or not yet popularized in their respective domains.<sup>13</sup>

In so doing, we will avoid giving an exceedingly normative outlook to our repertoire—as this would be incompatible with a linguistic use in evolution—but will not fail to provide a coherent and realistic picture of current developments.

<sup>12</sup> One could debate at length on the need of the average user to know the word *ماتلی* *matlî* ‘Martini-type rifle’ reported in S. Qazzaz (2000), despite this term has undoubtedly an interesting historical background (relating to the Ottoman Turkish-American production of rifles similar to those used by the British army, widely employed during the First World War in the Mesopotamian region).

<sup>13</sup> A criterion could be excluding any element attested in a number of occurrences inferior to a limit to define, with possible exceptions applicable to very specialized vocabularies or recent lexical innovations.



### 3.4. Orthography and Transcriptions

As already mentioned, Kurdish shows a regionally and dialectologically defined digraphia. An extended Latin script (having supplanted the use of the Cyrillic and Armenian alphabets) is used throughout the area of diffusion of Kurmanji varieties, with the partial exception of northern regions of Iraqi Kurdistan where Bahdini dialects are spoken, for which the use of the Arabic-based orthography is frequent. Also, several Latin-based scientific transcriptions, diverging to a certain extent from one another, are adopted in descriptive grammars of Central Kurdish and their respective lexicons and glossaries.

More recently, the informal network *Kurdish Academy of Language* (KAL) has advanced the suggestion to unify all Kurdish graphic norms under a simplified Latin-based system (according to the ISO-8859-1 standard)<sup>14</sup> named *yekgirtú*, *yekgirtí* or *yekgirig* ‘unified,’ characterized by a reduced number of diacritics. Although the initiative has been welcomed in the intentions, it is not yet possible to consider the graphic system as popularized among Sorani Kurdish writers and speakers, at least from the observation of linguistic uses on the Internet. Furthermore, this orthography does not appear in any of the recent dictionaries of Sorani.

About the dictionary compiler’s troubles in the identification of a uniform Latin transcription for Sorani, Kendal Nezan (2017: XXIV-XXV) writes: “L’ introduction du vocabulaire sorani dans un dictionnaire à base kurmandji a suscité diverses difficultés dont la première est l’écriture du sorani en caractères latins. À l’exception de *A Kurdish-English Dictionary* de Tawfiq Wahby et C. J. Edmonds, tous les dictionnaires soranis auxquels nous nous sommes référés sont écrits en caractères arabo-kurdes dont la transcription en caractères latins n’est pas fixée, ce qui peut parfois entraîner des confusions.”

The table that follows gives an idea of the extant variation in the orthographic and transcribing choices made by the compilers of the two most recent dictionaries of Sorani (S. Qazzaz, 2000; H. Hakem, 2012). Both of them report the headword in the modified Arabic script as primary, and secondly add a Latin-based transcription broadly akin to the Hawar system, except for a few differences highlighted below.

<sup>14</sup> This model has also been adopted in other attempts at graphic unification of Iranian languages with a predominantly oral tradition. See, for instance, the case of Balochi, for which the use of a similar system has become popular among native speakers on the web.

Phonetic value	Arabic-based Sorani script	Variants	“Hawar” script	Variants
[a:]	ا		A, a	â (Hakem)
[b]	ب		B, b	
[p]	پ		P, p	
[t]	ت		T, t	
[dʒ]	ج		C, c	
[tʃ]	چ		Ç, ç	
[h]	ح		(Ĥ, ĥ)	ĥ (Qazzaz); ĥ (Hakem)
[x]	خ		X, x	
[d]	د		D, d	
[r]	ر		R, r	
[ʀ]	ر		(Ř, ř)	ř (Qazzaz); ř (Hakem)
[z]	ز		Z, z	
[ʒ]	ژ		J, j	
[s]	س		S, s	
[ʃ]	ش		Ş, ş	ş (Qazzaz); ş (Hakem)
[ʕ]	ع		(ʕ)	ç (Qazzaz)
[y]	ع		(Û, ü)	Û (Qazzaz); ŭ (Hakem)
[f]	ف		F, f	
[v]	ف		V, v	
[q]	ق		Q, q	
[k]	ک	ك word-finally (Hakem)	K, k	
[g]	گ		G, g	
[l]	ل		L, l	
[ɫ]	ل		(Ľ, ľ)	ľ (Qazzaz); ľ (Hakem)
[m]	م		M, m	
[n]	ن		N, n	
[ɲ]	نگ		(Ñ, ñ)	ɲ (Qazzaz; Hakem)
[u]	و		U, u	
[w]	و		W, w	
[o]	و		O, o	
[u:]	وو		Û, û	
[∅]	وئ		—	
[ɛ]	ه		E, e	a (Hakem)
[h]	ه		H, h	
[e:]	ئ		Ê, ê	
[i:]	ی		Î, î	
[j]	ی		Y, y	
[ʔ]	ئ		—	' (Hakem)
[i]	—		I, i	
[y:]	ؤ	— (Qazzaz; Hakem)	(Û, ü)	— (Qazzaz; Hakem)

While the use of the modified Arabic script already standardized for Sorani seems difficult to question, the choice of an additional Latin script needs to be carefully evaluated and discussed. The chosen Latin orthography may either be entirely based on the Hawar system, or partly adapted to the scientific transcription used for Kurdish and other Iranian languages with Arabic-based scripts (Persian, Balochi, etc.).<sup>15</sup>

Furthermore, it will be necessary to assess the inclusion of IPA (International Phonetic Alphabet) transcriptions in the entries. They might be useful to highlight accent structures—even if these are overall predictable and uniform for Kurdish, where lexical stress typically falls on the last syllable—or draw the attention to phonetic and phonological phenomena that would otherwise remain non-transparent (e.g., / ng / [ŋ] in مانگ *mang* /ma:ŋ/ 1. *n.* luna, mezzaluna 2. *n.* mese ...).

### 3.5. Structure, Compilation and Ordering of Headwords

The alphabetical ordering of entries will depend on the choice of the Arabic-based (or rather Latin-based) orthography as primary in lemmatization. In the case of the Arabic-based script, some details—concerning, for instance, the ordering of vowels—will have to be determined in advance.

Homographs, identified by a superscript number, should be listed consecutively (as in K. Nezan and S. Qazzaz, but unlike H. Hakem who does not number them). The order will reflect a criterion of decreasing importance and frequency in common use, as illustrated in the following example:

- <sup>1</sup> سوور *sûr* /su:t/ 1. *agg.*, *n.* rosso (colore) ... 2. *agg.* ESTENS.FIG. risoluto, determinato ...  
<sup>2</sup> سوور *sûr* *n.* festa, cerimonia (spec. di nozze) ...  
<sup>3</sup> سوور *sûr* *n.* AGR. vaglio, spulatura (grano), separazione del cereale dai residui di trebbiatura ...

The same criterion will be applied to the ordering of meanings in the subsections of polysemous words. This process will require a careful and in-depth semantic analysis supported also, if needed, by etymological research. The meanings will be appropriately identified by progressive numbers in the architecture of each entry.

The choice of Italian glosses should aim at effectiveness, favouring simple equivalents representative of all main senses. Whenever the attribution of certain senses to one or more simple Italian words proves impossible, recourse can be made to short phrases or definitions. These will be of considerable importance also for items pertaining to scientific, technical or other specialized vocabularies.

At least one grammatical function must be indicated for each entry. The grammatical marks may be more than one if certain meanings possess distinct qualifications or several possible qualifications. The grammatical marks will also be accompanied by one or more usage notes as in H. Hakem (2017), who for the first time systematically introduced this praxis in a Kurdish dictionary. Indeed, the inclusion of usage notes is still overall uncommon

<sup>15</sup> Among the available choices, is the use of <č> for [tʃ]; <ǰ> or <ǰ> for [dʒ]; <ǰ> for [ʒ]; <ğ> or <ɣ> for [ɣ]; <ř> (id. Hawar) for [r]; <ǎ> or <ǎ> (id. Hawar) for [ɬ].

in lexicographic works, but proves extremely handy for the reader who obtains from them fundamental indications on the usage of words in specific domains, styles, or registers.

Given the very high frequency of complex verbs with nominal or particle components, these will need to be included within the entries relating to their non-verbal constituents, under the section corresponding more closely to the semantics of the resulting complex verb. Simple verbs, not as numerous as complex ones, will be lemmatized separately in the Infinitive. In this case, the Present stem will also be reported, considering that it is often unpredictable and not immediately recoverable from the Infinitive form.

Any idiomatic phrase or expression may also appear under the entries corresponding to their constituent words. The inclusion of additional sections relating to COMPOUNDS, DERIVATES, VARIANTS, GRAMMAR, etc.<sup>16</sup> will be carefully assessed: their structure and scope should be defined in detail, taking into account the need to ensure a certain uniformity of entries, easy retrieval of information, and a final product of reasonable dimensions. Coherent and non-extensive use of cross-references may also serve the purpose.

Depending on the feasibility of a systematic analysis of textual and oral materials and the availability of a balanced language sample, phraseologies will be identified and inserted to disambiguate or illustrate particular uses, when needed.

### 3.6. Use of IT Tools and Lexicographic Software

Since the earliest stages of the work, it will be necessary to scrutinize all multilingual lexicographic software available on the market to assess their functionalities and potential for the automatic completion of selected tasks, as this could shorten considerably the time needed for the final processing of the product. This phase may require the advice of experts in the IT sector.

Relevant for our purposes are also some digitalization projects concerning monolingual and bilingual Kurdish dictionaries – e.g., the *Vejin Dictionaries: Kurdish Dictionaries Collection* (<https://lex.vejin.net/en/>)—as well as other printed materials—cf. the *Kurdish Digital Library* (BNK) of the Institut kurde de Paris (<https://bnk.institutkurde.org/?l=en>).

Moreover, some information about recent projects aimed at the elaboration of balanced corpora for Sorani Kurdish are retrievable online: among these, the *AsoSoft text corpus*<sup>17</sup> (partly available at <http://asosoft.com/>) and the corpus announced by *Sketch Engine*<sup>18</sup> (<https://www.sketchengine.eu/kurdish-sorani-text-corpora/>). The *Dialects of Kurdish* database (Y. Matras et al., 2016, University of Manchester. <http://kurdish.humanities.manchester.ac.uk/>) is also available on the Internet, although this has been developed mostly in the scope of comparative dialectology and, given its high specialization, is only partially useful for a lexicographic collection as the one discussed here.

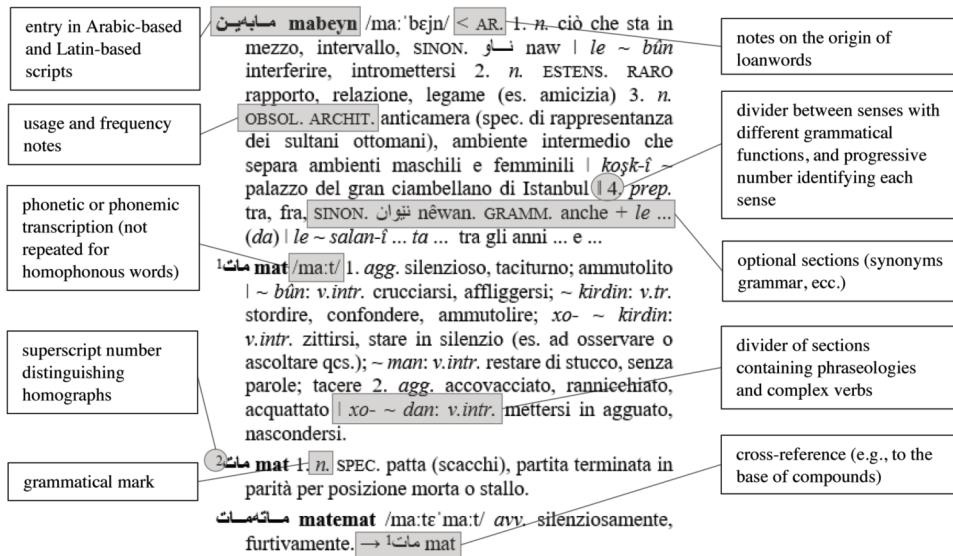
<sup>16</sup> Akin to those in the *Grande dizionario italiano dell'uso* (GRADIT) edited by Tullio de Mauro (2<sup>nd</sup> ed., 8 vols., 2007, Turin: Utet).

<sup>17</sup> Si veda in proposito M. Mohammadamini, H. Veisi & H. Hosseini, 2019, “Toward Kurdish language processing: Experiments in collecting and processing the AsoSoft text corpus,” *Digital Scholarship in the Humanities* 35 (1), 176-193.

<sup>18</sup> *Sketch Engine* is presented online as “the ultimate tool to explore how language works. Its algorithms analyze authentic texts of billions of words (text corpora) to identify instantly what is typical in language and what is rare, unusual or emerging usage.” This feature might prove extremely useful in the case of Sorani.

Almost two decades have passed since the publication of the *Dictionnaire français-kurde* (H. Hakem, G. Gautier, 1993, Paris: Klincksieck) and the technological tools available to lexicographers are now much more advanced than in the past. Even so, it is still worth considering the discussion on the problems and possible solutions for computerized data processing in G. Gautier, 1996, *Dirêjî Kurdî* درێژی کوردی: *A lexicographic environment for Kurdish language using 4<sup>th</sup> Dimension®*, 5<sup>th</sup> International Conference and Exhibition on Multilingual Computing (ICEMCO), London.

#### 4. Un modello preliminare del Dizionario bilingue Kurdo (Sorani) – Italiano





layout example →



## م M, m

م **M, m** lettera dell'alfabeto kurdo-sorani denominata *mîm* /mi:m/.

م **-im** /im/ 1. *pro. suff.* 1ª pers. sg. | 2. *suff. desinenza* 1ª pers. sg. presente e passato | 3. *v.tr.* presente indicativo 1ª pers. sg. del verbo 'essere' → بوون *bûn* | GRAMM. -m dopo vocale.

1ما **ma-** /ma:/ *pref.* ZOOLOG. femmina (animale), SINON. دله *dele-* (spec. canidi), CONTR. نيره *nêre-* | ~ker asina; ~kew femmina di pernice. GRAMM. corrisponde a م *mê* posto dopo il nome in rapporto d'annessione.

2ما **ma** *n.* RARO montagna (spec. aspra, scoscesa). SINON. کيو *kêw*.

3ما **ma** < AR., RARO *avv.* perché mai? SINON. بوچی *boçî*.

مابهين **mabeyn** /ma:'bejn/ < AR. 1. *n.* ciò che sta in mezzo, intervallo, SINON. ناون *naw* | *le* ~ *bûn* interferire, intromettersi 2. *n.* ESTENS. RARO rapporto, relazione, legame (es. amicizia) 3. *n.* OBSOL. ARCHIT. anticamera (spec. di rappresentanza dei sultani ottomani), ambiente intermedio che separa ambienti maschili e femminili | *koşk-î* ~ palazzo del gran ciambellano di Istanbul | 4. *prep.* tra, fra, SINON. نوان *nêwan*. GRAMM. anche + *le* ... (*da*) | *le* ~ *salan-î* ... *ta* ... tra gli anni ... e ...

1مات **mat** /ma:t/ 1. *agg.* silenzioso, taciturno; ammutolito | ~ *bûn*: *v.intr.* cruciarsi, affliggersi; ~ *kirdin*: *v.tr.* stordire, confondere, ammutolire; *xo-* ~ *kirdin*: *v.intr.* zittirsi, stare in silenzio (es. ad osservare o ascoltare qcs.); ~ *man*: *v.intr.* restare di stucco, senza parole; tacere 2. *agg.* accovacciato, rannicchiato, acquattato | *xo-* ~ *dan*: *v.intr.* mettersi in agguato, nascondersi.

2مات **mat** 1. *n.* SPEC. patta (scacchi), partita terminata in parità per posizione morta o stallo.

ماتاو **mataw** /ma:'taw/ *n.* REGION. chiaro di luna, SINON. مانگهشو *mangêşew*.

ماتک **matik** /ma:'tik/ < PERS.? 1. *n.* mascara 2. *n.* RARO kajal, matita per occhi, SINON. کل *kil*.

ماتلی **matîlî** /ma:'tî:/ *n.* OBSOL. MILIT. fucile di tipo Martini, usato spec. in epoca ottomana.

ماتماتیک **matmatîk** /matma:'ti:k/ 1. *n.* matematica (disciplina), SINON. بیرکاری *bîrkarî* | 2. *agg.* (di qcs.) matematico.

ماتماتیکدان **matmatîkdan** /mat,ma:'ti:k'da:n/ *n.* VAR. → ماتماتیکزان *matmatîkzan*

ماتماتیکزان **matmatîkzan** /mat,ma:'ti:k'za:n/ 1. *n.* (di qcn.) matematico, -a (professione) | 2. *agg.* esperto in matematica. → ماتماتیک *matmatîk*

ماتماتیکى **matmatîkî** /mat,ma:'ti:'ki:/ *adj.* (di qcs.) relativo alla matematica, matematico. → ماتماتیک *matmatîk*

ماتور **mator** /ma:'to:/ *n.* VAR. → ماتور *matoř*

ماتورسایکیل **matorsaykil** /ma:'torsaj'kil/ *n.* VAR. → ماتورسایکیل *matořsaykil*

ماتوره **matore** /ma:'to:re/ *agg.* RARO (spec. di qcs.) insignificante, futile, vile, SINON. پوپوچ *piřupuç*.

ماتور **matoř** /ma:'to:ř/ < FR.-INGL 1. *n.* motore 2. *n.* FAM. moto, ACCORC. → ماتورسایکیل *matořsaykil*

ماتورسایکیل **matořsaykil** /ma:'torsaj'kil/ < INGL. *n.* motociclo, motocicletta.

ماتورسوار **matořsiwar** /ma:'to:řsi'wa:t/ *n.* motociclista. → ماتور *mator*

ماتورسوارى **matořsiwarî** /ma:'to:řsi'wa:ri:/ *n.* motociclismo. → ماتورسوار *matořsiwar*

ماته **mate** /ma:'te/ 1. *agg.* RARO verde, acerbo, non del tutto maturo (frutta e verdura), SINON. فەرىک *ferîk* | spec. in ~ *dan*: *v.intr.* essere in maturazione, maturare.

ماتەریال **materyal** /ma:'te:'rja:l/ < FR. 1. *n.* materiale, SINON. کەرستە *kereste* | 2. *agg.* (di qcs.) materiale.

ماتەریالی **materyalî** /ma:'te:rja:'li:/ 1. *n.* materialità | 2. *agg.* (di qcs.) relativo al materiale o alla materialità; ESTENS. FILOS. materiale (opposto a spirituale), CONTR. روحى *ruhî*. → ماتەریال *materyal*

ماتەریالیست **materyalist** /ma:'te:rja:'li:st/ < FR.-INGL. 1. *agg.* FILOS. (di qcn.) materialista, SINON. پەزەوار *pêzewar*; *madepêrist*; مادەگیرا *madegira*. → ماتەریال *materyal*

ماتەریالیزم **materyalizm** /ma:'te:rja:'li:zm/ < FR. *n.* FILOS. materialismo, SINON. پەزەوارى *pêzewarî*; *madepêristî*; مادەگیرایى *madegirayî*; مادىبەت *madîyêt*. → ماتەریال *materyal*

ماتم **matem** /ma:'tem/ < AR., LETT. *n.* lutto, cordoglio | ~ *girtin*: osservare il lutto, essere in lutto, SINON. پرسە *pirse*; شین *şîn*; شیوان *şîwen*.

ماتمات **matemat** /ma:'te'ma:t/ *avv.* silenziosamente, furtivamente. → ماتا *mat*

ماتماتە **matemate** /ma:'tema:'te/ 1. *n.* movimento furtivo | *be* ~: *avv.* in maniera circospetta o guardinga, furtivamente. → ماتا *mat*

ماتمى **matemî** /ma:'te'mi:/ *n.* VAR. → ماتم *matem*

ماتمەين **matemîn** /ma:'te'mi:n/ 1. *agg.* (spec. di qcn.) in lutto; ESTENS. GENER. afflitto, sconcolato 2. *agg.* (di qcs.) di lutto, luttuoso. → ماتم *matem*

ماتمەينى **matemîni** /ma:'temi:'ni:/ 1. *n.* pena, afflizione, sconfitta 2. *n.* condoglianze | *bîruske-î* ~ telegramma di condoglianze. → ماتمەين *matemîn*

ماتى **matî** /ma:'ti:/ 1. *n.* silenzio (spec. assorto, penseroso o sognante); malinconia 2. *n.* MEDIC. depressione, melanconia. → ماتا *mat*

ماچ **maç** /ma:'tʃ/ *n.* bacio | ~ *kirdin*: *v.tr.* baciare (qcn. o qcs.), *em la-u ew la* ~ *kirdin* scambiarsi baci in segno di saluto; *yektirî* ~ *kirdin* baciarsi, scambiare effusioni; IDIOM. *dest* ~ *kirdin* fare il baciamano, FIG. supplicare, implorare; *qondere* ~ *kirdin* leccare i piedi a qcn., *adulare* servilmente per ottenere un vantaggio (lett. baciare le scarpe); ~ *dan*: *v.intr.* dare un bacio (a qcn.).

ماچباران **maçbaran** /ma:'tʃba:'ra:n/ *n.* baci a profusione (lett. pioggia di baci). → ماچ *maç*; باران *baran*

## AN ESTIMATED BUDGET (IN EURO)

Project Director (3 years × 60,000 per year)	180.000
Project assistants (2 years × 30,000 × 2)	120.000
Office overheads	80.000
Native-speaker informants	80.000
Pre- and publishing costs, communication and dissemination activities, etc. (1000/2000 copies)	80.000
tot.	<u>540.000</u>