Regional Studies Journal
https://regs.uomosul.edu.iq/

Language Distribution in Al-Hamdaniyah District, Ninawa Province, Iraq

***Hakam Ghanim*[1]** ***Erik Anonby*[2]**

[1] *Assist. Lect. / School of linguistics and language studies/ Carleton University/ Canada.*
hakamghanim@cmail.carleton.ca
[2]*Prof./ Linguistics and French Dep./ Carleton University/ Canada.*
erik.anonby@carleton.ca

Article Information
Received: 27/4/2026
Revised: 2/6/2026
Accepted: 25/6/2026
Published 1/7/2026

Corresponding

Hakam Ghanim

Keywords

Language Mapping; Linguistic Geography; Language Distribution; Iraq; Ninawa Province; al-Hamdaniyah District.

Citation

Ghanim, H., & Anonby, E. (2026). Language Distribution in Al-Hamdaniyah District, Ninawa Province, Iraq. Regional Studies Journal. 20 (69). 423–460. https://doi.org/10.33899/rsj.v20i69.62785

Abstract

Iraq is a linguistically diverse country where languages from several language families are spoken. Prolonged political instability has led to internal displacement and migration, making it more difficult to understand the spatial distribution of these languages. In many parts of Iraq, linguistic documentation remains limited, whether in the form of language maps or other language-related studies.

This study addresses this gap through the systematic documentation of the spatial distribution of linguistic communities in Al-Hamdaniyah District, Nineveh Province, northern Iraq. While earlier language maps relied largely on general knowledge and sources such as newspaper reports, this study combines specialized demographic and geographic records with firsthand fieldwork conducted across settlements in the district.

For data visualization, the open-source GIS software QGIS was employed. The study produced a map showing the distribution of languages in Al-Hamdaniyah District. This map contributes to a better understanding of the region's linguistic landscape, gives greater visibility to underrepresented linguistic communities, and assists in selecting representative locations for future comparative research on linguistic features using a language-data questionnaire.

مجلة دراسات إقليمية
https://regs.uomosul.edu.iq/

توزيع اللغة في منطقة الحمدانية، محافظة نينوى، العراق

حكم غانم[١] إريك أنونبي[٢]

[١] مدرس مساعد/ قسم علم اللغة والدراسات اللغوية/ جامعة كارلتون. كندا. hakamghanim@cmail.carleton.ca

[٢] أستاذ/ قسم علم اللغة واللغة الفرنسية/ جامعة كارلتون. كندا. erik.anonby@carleton.ca

معلومات الأرشفة

الاستلام: ٢٠٢٦/٤/٢٧
المراجعة: ٢٠٢٦/٦/٢
القبول: ٢٠٢٦/٦/٢٥
النشر الإلكتروني: ٢٠٢٦/٧/١

المراسلة
حكم غانم

الكلمات المفتاحية

رسم خرائط اللغات؛ الجغرافيا اللغوية؛ توزيع اللغة؛ العراق؛ محافظة نينوى؛ منطقة الحمدانية.

الاقتباس

غانم، حكم.، وأنونبي، إ. (٢٠٢٦) توزيع اللغة في منطقة الحمدانية، محافظة نينوى، العراق. مجلة دراسات إقليمية. ٢٠ (٦٩). ٤٢٣-٤٦٠.
https://doi.org/10.33899/rsj.v20i69.62785

الملخص

العراق بلد متنوع لغويا حيث تتحدث لغات مختلفة من عائلات لغوية مختلفة. كما أنها منطقة من عدم الاستقرار السياسي، مما يؤدي إلى نزوح داخلي وهجرة الناس، مما يعقد مهمة فهم التوزيع المكاني لهذه اللغات. في بعض مناطق العراق، لم يكن هناك توثيق لغوي كبير، سواء كان هناك خرائط أو أبحاث لغوية أخرى. في هذه الدراسة، نسعى لمعالجة هذه الفجوة من خلال توثيق منهجي للتوزيع المكاني لمجتمعات اللغة في منطقة الحمدانية، محافظة نينوة، شمال العراق. بينما اعتمدت خرائط اللغات السابقة على المعرفة العامة ومصادر مثل مقالات الصحف، فإن المنهجية المستخدمة في هذه الدراسة تتضمن مزيجا من الوثائق الديموغرافية والجغرافية المتخصصة والعمل الميداني المباشر في المستوطنات عبر المنطقة. لتمثيل البيانات، تم استخدام برنامج خرائط اللغة مفتوح المصدر QGIS. نتيجة هذه الدراسة هي خريطة تظهر توزيع اللغة في منطقة الحمدانية. ستساهم هذه الخريطة في فهم أكثر تفصيلا لوضع اللغة، والتفكير بشكل أفضل في المجتمعات اللغوية الممثلة تمثيلا ناقصا، وتوجيه اختيار المواقع الممثلة للبحث المقارن لاحقا للخصائص اللغوية باستخدام استبيان بيانات اللغة.

1. Introduction

Iraq is a linguistically diverse country. Many languages are spoken there, and most are part of three major language families: Semitic (Arabic and Neo-Aramaic), Iranic (Kurdish, Gorani, Shabaki, Kakai), and Turkic (Turkmen). The linguistic situation in Iraq has been unstable in recent decades due to political factors. Many ethnic groups that speak different languages have been displaced over the years. Most of these communities or groups are religious minorities, such as Christians, Jews, and Ayzidis (Khan & Haig, 2019).

This paper investigates language distribution in Iraq, specifically in the northern part of Iraq, in a district within Ninawa Province known as al-Hamdaniyah (Ar. al-Ḥamdānīyah). First and foremost, it is useful to discuss the classifications of languages in al-Hamdaniyah before looking at their distribution. The Semitic language group is represented by Arabic and Neo-Aramaic. The Arabic spoken there has two main divisions: *qeltu* and *gelet* (Blanc, 1964; Palva, 2009). Speakers of *qeltu* varieties are mainly located in the northern part of the country, while speakers of *gelet* varieties are mainly located farther south and in the suburbs of Ninawa (Blanc, 1964; Prochazka, 2019). Within *qeltu*, Moslawi is the most widely spoken dialect and is mainly spoken in the city of Mosul (Ar. al-Mawṣil) in Ninawa Province (Jastrow, 2006; Prochazka, 2019). Neo-Aramaic is represented by varieties belonging to a group of dialects known as North-Eastern Neo-Aramaic (NENA). One of the NENA dialects is Surith (also called Surayi) and is spoken in al-Hamdaniyah, and it is also known as Ashuri or Athuri (Assyrian) or Syriac (Khan, 2007; Khan, 2016). The writing system of the NENA dialect group is known as Syriac (Khan, 2007). The Iranic languages in Ninawa are represented by Kurdish and its major subgroups, Northern Kurdish (Kurmanji or Bahdini) and Central Kurdish (Sorani), and by the Gorani group, which is represented by Shabaki, Bajelani, and Kakai (Haig & Öpengin, 2018). These language groups fall under a branch of the Western division of Iranic languages within the Indo-European language family and are spoken by Kurds, Shabaks, and Kakais, who primarily inhabit regions in Iran, Iraq, Syria, and Turkey (McCarus, 2009; Paul, 2008). In Iraq, Kurds constitute a significant portion of the population and are mainly concentrated in the autonomous Kurdistan Region (KRG) and in parts of Ninawa Province (McDowall, 2004; Natali, 2005). According to Postgate (2007), the distribution of the languages in Iraq mirrors the natural geographical areas there, with Semitic languages generally found in lower areas, and Iranic languages concentrated in mountainous regions, as well as Turkic represented by Turkmen in some areas.

The language situation in Iraq is complex, particularly in the northern part where minorities such as Shabaki, Kakai, Turkmen, and Assyrians live. Moreover, existing research on this complex language situation is limited, since

there are not enough studies that look at it, nor are there maps that reflect the distribution of its languages, except for a few general maps that will be discussed later in this paper. There is also a lack of language documentation for the minority languages spoken in Iraq. This paper seeks to address this gap by looking at the distribution of languages in al-Hamdaniyah District to spatially document the languages spoken there.

Al-Hamdaniyah District was chosen for this study because the language situation there is diverse. This study was conducted at the settlement level to understand the distribution of languages in Ninawa. This is the first detailed and comprehensive study on the distribution of languages in Ninawa in general and in al-Hamdaniyah in particular.

This paper is organized as follows. Section 2 presents the geographical locations of Ninawa Province and al-Hamdaniyah District. Section 3 presents maps that visualize the distribution of languages in Ninawa. Section 4 presents the research process and methodology as it looks at data collection, the incorporation of existing data, other documentation efforts, and the collection of new language distribution data. Section 5 shows the results on language distribution in al-Hamdaniyah. Section 6 offers a discussion of the results, while Section 7 is the conclusion.

2. Geographical Locations

This section provides the geographical context for the study by outlining the physical setting, administrative boundaries, and settlement pattern of Ninawa Province, particularly al-Hamdaniyah District. It situates the province within northern Iraq and highlights its internal district and subdistrict organization. Additionally, it identifies the village network that serves as the unit of analysis for the language distribution map.

Ninawa Province

Ninawa Province is in the northwest of Iraq. It borders the Kurdistan Region of Iraq to the north and the Syrian border to the west. This province holds a great significance within the historical, cultural, and geographical context of the country. It has a rich history and diverse demographics. Additionally, Ninawa Province has a blend of ethnic and religious communities, as well as a deep historical legacy, with roots extending back to ancient times. This region is known for being home to the ancient city of “Nineveh” the capital of the Assyrian Empire. The Tigris River runs through this province, dividing it into two parts. The eastern part includes the districts of Tal Kef “Tal Keppe”, al-Hamdaniyah, al-Shekhan, Makhmour, and parts of Mosul District, while the western part includes parts of Mosul and the districts of Sinjar, Tal Afar, al-Baaj, and al-Hadhar. The population of Ninawa Province is between 4.5 and 5 million people, according to the preliminary results of the most recent census of 2024. This

province has an ethnically, linguistically, and religiously diverse population, which makes it stand out as a microcosm of Iraq. Ninawa is home to many ethnic groups, especially Arabs, Assyrians, Chaldeans, Kurds, Turkmen, Ayzidis, Shabakis, and Kakais. Most of the languages of Iraq are spoken here as well. There have been no comprehensive studies on this region, and there are not many studies that aim to fill this gap and describe the linguistic diversity of Ninawa Province. Ninawa Province has undergone the most demographic changes of any area in Iraq due to political instability. Even though this Province has faced many challenges, including conflict, displacement, and the rise of extremist groups in the past and in recent times, it remains a melting pot of cultures, traditions, and histories, and it is the most linguistically diverse region in Iraq. Figure 1 below presents an administrative map of Ninawa Province.[1]

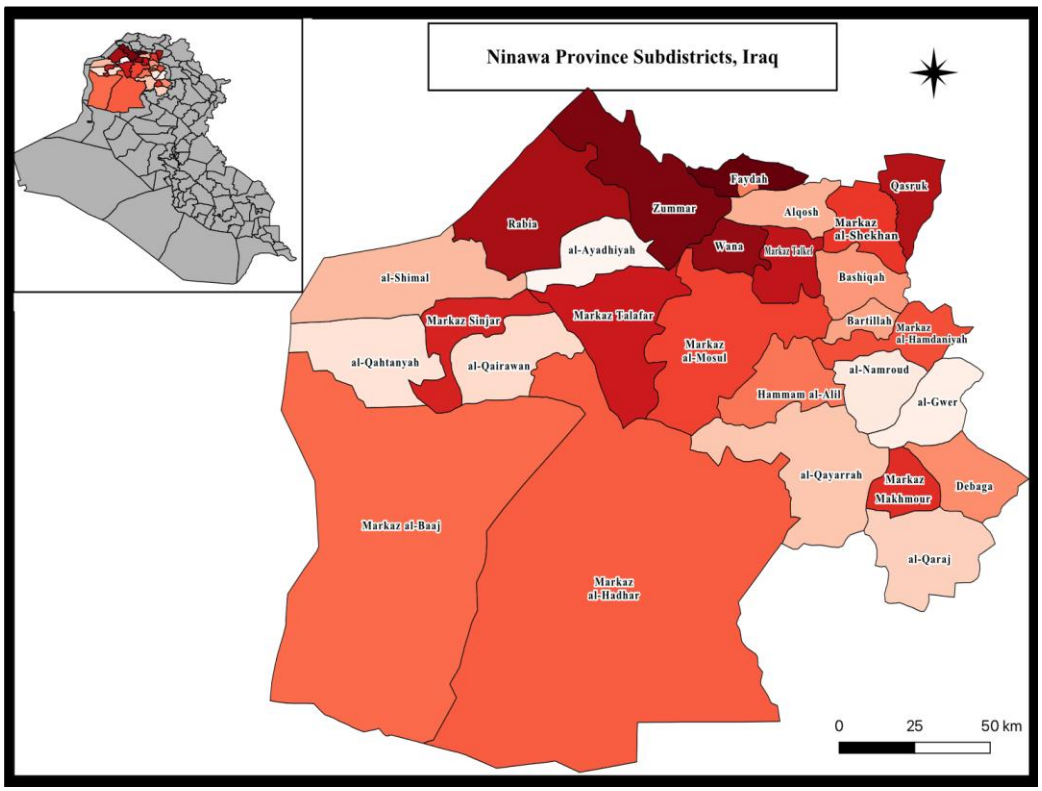

Figure 1. Ninawa Province and its districts and subdistricts (Map H. Ghanim, 2025. Data: HDX Iraq shape files. Produced in QGIS.)

[1] Created by Hakam Ghanim using QGIS with a shapefile layer from Humanitarian Data Exchange https://data.humdata.org/dataset/cod-ab-irq.

Al-Hamdaniyah District

As seen in the map in Figure 1 above, al-Hamdaniyah is one of nine districts in Ninawa Province. This district is also known as Baghdida (Ar. Baghdīda) by the Christian population living there. There are three subdistricts in al-Hamdaniyah: The Centre of al-Hamdaniyah (Ar. Markaz al-Ḥamdānīyah), al-Namrud (Ar. al-Namrūd), and Bartilla (Ar. Barṭilla). There are 131 settlements (cities, towns, and villages) presented in the fieldwork data for the language distribution map.[1] This district is the most linguistically diverse district in Ninawa as it is inhabited by different ethnoreligious groups such as Muslims, who are mostly Arabs, Kurds, Shabaks, Kakais, and Turkmen; Christians, who can be Arabs or Kurds; and Ayzidis. Each one of these groups speaks a characteristic language. Arabic is the dominant language of the province and serves as a lingua franca. To the west and northwest of al-Hamdaniyah lies Mosul District, where Arabic is spoken as the primary language in its two varieties, *qeltu* and *gelet*. Al-Shekhan District lies to the north, where Northern Kurdish (Bahdini or Kurmanji) and *gelet* Arabic are spoken. Aqre District lies to the northeast, where Northern and Central Kurdish (Bahdini/Kurmanji and Sorani) are mainly spoken, and Makhmour (Ar. Makhmūr) lies to the southeast, where Central Kurdish (Sorani) and *gelet* Arabic are mainly spoken. This district is also part of what is known as Ninawa Plain (Ar. Sahil Naynawa), which includes Talkef and al-Hamdaniyah Districts.[2]

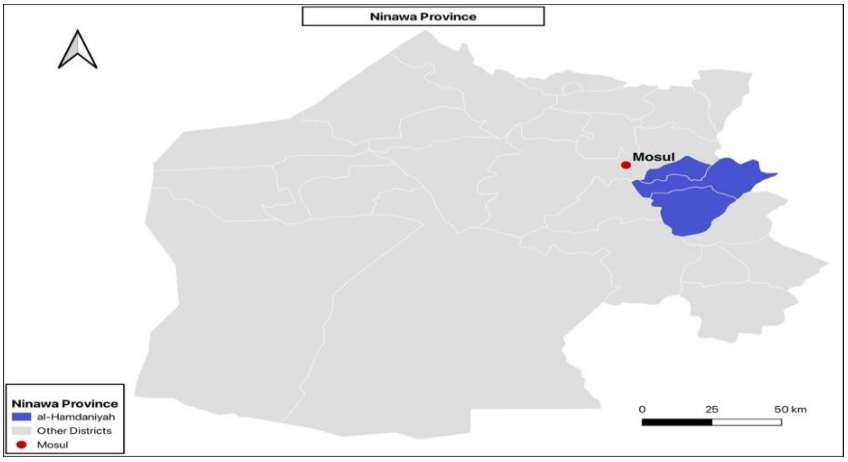

Figure 2. Background map of Ninawa Province showing the geographical location of al-Hamdaniyah District (Map by H. Ghanim, 2025. Data: HDX Iraq shape files. Produced in QGIS.)

[1] There might be more villages as the 131 represents the number of populated places found in geographical sources and during field work. The term 'settlements' is used here to refer to any populated place. It can refer to cities, towns, and villages.

[2] Produced by Hakam Ghanim using QGIS with a shapefile layer from Humanitarian Data Exchange https://data.humdata.org/dataset/cod-ab-irq

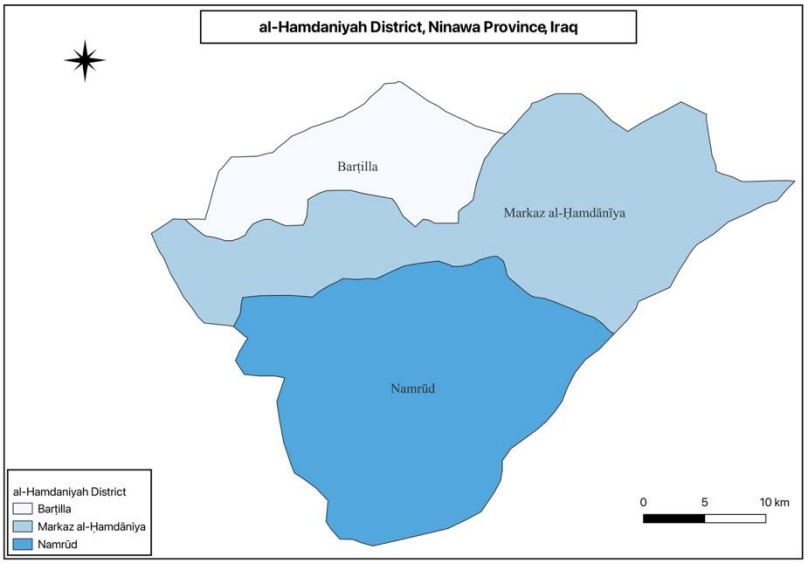

Figure 3. A map of al-Hamdaniyah district showing its three subdistricts (Map by H. Ghanim, 2025. Data: HDX Iraq shape files. Produced in QGIS.)

3. **Maps Featuring Languages of Ninawa Province**

There are several studies that provide an overview of language distribution in Iraq through maps. One of these maps is Izady's (2014) language map. It uses colour-coded polygons to indicate the locations of the languages spoken there. This map is part of the Gulf 2000 Project that is home to the *Atlas of the Islamic World and Vicinity*. It is the most comprehensive map produced thus far that focuses on the distribution of languages in Iraq. However, it still does not give a full picture of the language distribution; although the data collection was comprehensive in its coverage, it was based largely on general knowledge about the country from news and political journals (Ghanim, 2026).[1]

[1] This is from an online interview with the author of this map M. M. Izady in 2021. The interview was conducted as part of another study that looks into the technicalities of the map (Ghanim, 2026).

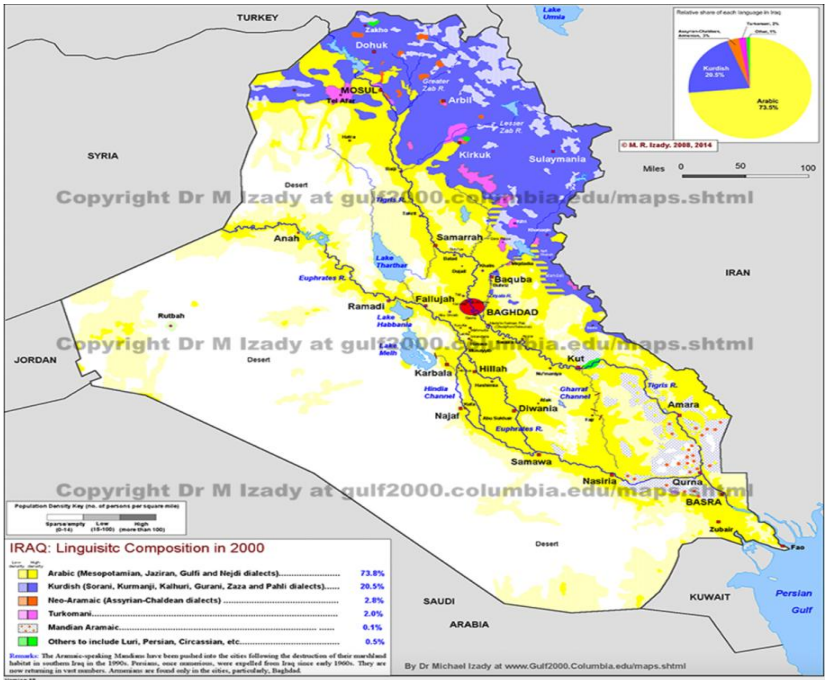

Figure 4. Language Distribution in Iraq (Source: Dr M. Izady, Atlas of the Islamic World and Vicinity (New York, Columbia Univ., 2006-present), @gulf2000.columbia.edu/maps.shtml, used with permission)

The map by Haig and Khan (2019) shows the location of languages in Western Asia, including those spoken in the northern part of Iraq. As shown in the map, Arabic, which is from the Semitic group, is spoken in the north of Iraq. It also shows that Gorani and Turkmen are spoken in parts of Ninawa. The map is not very comprehensive when it comes to the distribution of the languages in the north of Iraq, since it pinpoints only one location where the groups of languages are spoken. It also does not show Aramaic, which is spoken in the northern part of Iraq.

Figure 5. Distribution of Languages in Western Asia (Reproduced from Haig and Khan, The Languages and Linguistics of Western Asia: An aerial perspective 2019: p. 2)

Khan (2019) provides another map that shows the distribution of North-Eastern Neo-Aramaic (NENA) varieties. This map groups the languages based on their shared linguistic features. With respect to al-Hamdaniyah, the map shows only that a NENA variety is spoken in the subdistricts of Barṭillah, Karamles, and Qaraqosh (the latter also known as al-Ḥamdānīyah City). This map does not comprehensively reflect the distribution of the languages in general or the NENA varieties in al-Hamdaniyah. It provides only a general view of where these languages or varieties are spoken across northern Iraq (Khan, 2019). In the legend of the map, IC1, IC2, and IC3 represent different varieties of Aramaic spoken in northern Iraq. IC1 represents the Christian dialects of the Mosul Plain in Ninawa Province, IC2 represents the dialects spoken up north close to the Turkish border in Duhok Province mainly west of the Zab, and IC3 represents the dialects east of the Zab in Erbil and al-Suleimaniyah Provinces. The variety spoken in al-Hamdaniyah is IC1, as shown on the map (Khan, 2019).

Figure 6. Distribution of Aramaic Languages in Northern Iraq (Reproduced from Khan, The Neo-Aramaic languages of Northern Iraq, 2019, p. 307-308)

Prochazka (2019) presents a point map that shows the distribution of Arabic in northern Iraq (Prochazka, 2019). It does not provide the whole picture of the distribution of Arabic in al-Hamdaniyah. It does not even show that Arabic is spoken there. However, as will be noted in the language-distribution map in this paper, Arabic is widely spoken in many settlements in al-Hamdaniyah district.

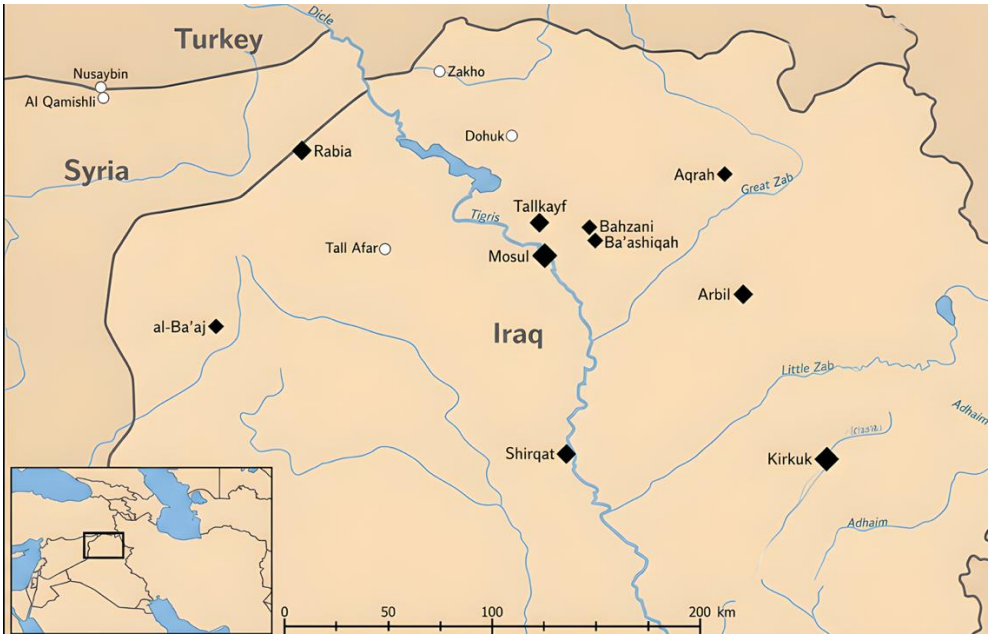

Figure 7. Distribution of Arabic dialects in the North of Iraq (Reproduced from Prochazka, The Arabic dialects of northern Iraq, 2019, p. 243)

Regarding Iranic languages in the north of Iraq, Allison (2007) presents a polygon map that shows the distribution of Kurdish and its varieties across Kurdistan, including the north of Iraq. The map shows that Northern Kurdish (Bahdini/Kurmanji) is spoken across al-Hamdaniyah District. However, the map does not show precisely in which settlements it is spoken or whether there are any other varieties spoken along with it.

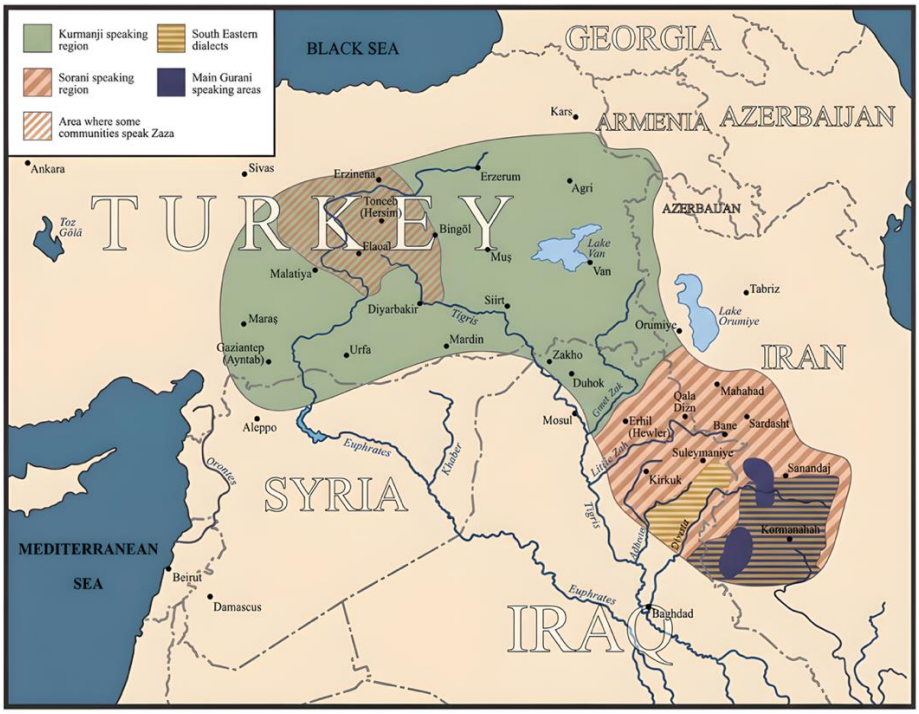

Figure 8. Distribution of Kurdish (Reproduced from Hassanpour A., Nationalism and Language in Kurdistan, 1992, p. 22.)

Gorani is another group within the Iranic languages with Shabaki, Bajilani, and Kakai. Haig (2019) provides a point map that shows where these Gorani varieties are spoken. The map shows that only Shabaki and Bajilani are spoken in al-Hamdaniyah district.

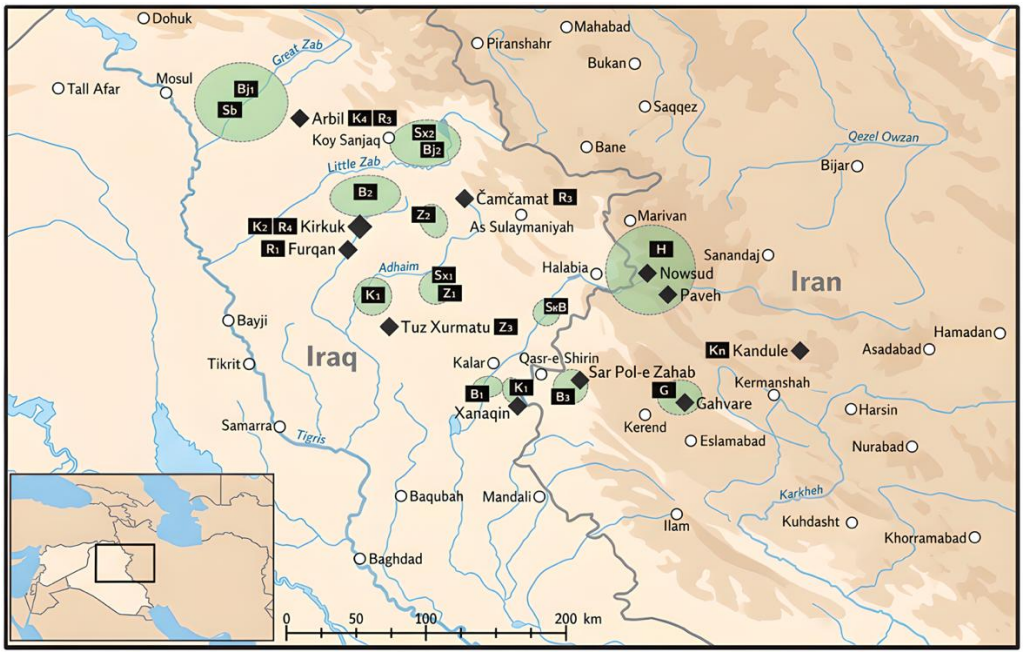

Figure 9. Distribution of Gorani (Shabaki, Bajilani, and Kakai) in Northern Iraq (Reproduced from Haig, G., The Iranian languages of northern Iraq, 2019, p. 297)

For Turkic, there were two maps that show the distribution of this language group in northern Iraq. For instance, Bulut (2007) presents a map of the distribution of Turkmen languages in Iraq. It shows the distribution of three different Turkmen varieties, including the y-group and the w-group, across Iraq. This map is more detailed than the other maps we have looked at thus far because it shows the settlements where Turkmen is spoken across Iraq. For example, it shows that Turkmen is only spoken in the Kalak settlement in al-Hamdaniyah District.

Figure 10. The Distribution of Turkmen in Iraq (Reproduced from Bulut, C. Iraqi Turkmen (2007), p. 160)

The previous maps provide an overview of language distribution in al-Hamdaniyah District. The maps in Figures 5 and 6 show that Aramaic is spoken in al-Hamdaniyah, specifically in Qaraqosh, Karamles, and Bartilla. The map in Figure 7 shows that Arabic is not spoken in al-Hamdaniyah District, which differs from what is found in the fieldwork. The maps in Figures 8 and 9 show that Northern Kurdish is spoken in part of al-Hamdaniyah District and show that Shabaki is spoken in small parts of the district. In contrast, our findings show that Shabaki is one of the most widespread languages in al-Hamdaniyah District. Finally, the map in Figure 10 shows that Turkmen is spoken in only one settlement in al-Hamdaniyah, whereas it can be found in more places around the district nowadays. The maps are useful in the sense that they can be built on as they provide a decent picture of language distribution that can be improved with recent data from fieldwork.

The maps reviewed above, Izady (2014), Haig and Khan (2019), Khan (2019), Prochazka (2019), Hassanpour (1992), Haig (2019), and Bulut (2007), provide a valuable point of departure for understanding language distribution in Iraq and in Ninawa in particular. At the same time, they share limitations that are especially salient for al-Hamdaniyah District, such as reliance on generalized sources, coarse polygon or single-point representations, omission of certain languages in some cases, and a lack of settlement-level detail. Most are not up-to-date, and their

methodologies do not primarily depend on fieldwork or on direct input from speakers and communities. These constraints help explain discrepancies with our findings (e.g., the extent of Arabic, Shabaki, and Turkmen in al-Hamdaniyah) and underscore the need for a more up-to-date, community-informed map. The next section, therefore, builds on this prior work by presenting a settlement-level, fieldwork-based mapping of al-Hamdaniyah that depends on input from speakers and local communities, including figures that estimate the proportions of languages in each settlement.

4. **Research Process and Methodology**

In this section, the research process of this study is described. The data collection methodology and the nature of the data collected are also discussed. Then, existing data sources on the language distribution in Ninawa and in al-Hamdaniyah district specifically are presented along with the new data for language distribution.

Data Collection

This study employed a descriptive survey methodology. Data were collected through field interviews using a structured questionnaire administered by the first author (H. Ghanim) to a sample of 107 participants. The survey approach was selected to gather firsthand information on the linguistic situation in al-Hamdaniyah District. The methodological framework of this study was implemented over a four-year period, from 2021 to 2024. All field data collection activities were carried out during this timeframe, which constitutes the temporal scope of the research. The fieldwork was conducted across settlements in al-Hamdaniyah District, Ninawa Province, and it aimed to systematically document the languages spoken in each settlement and the approximate proportions of speakers for each language.

Participant Recruitment and Selection

Participants were recruited through a goal-directed sampling strategy designed to identify knowledgeable individuals with deep familiarity with the linguistic composition of their communities. In total, 107 language consultants who represent each of the language communities in the province and possess knowledge of the languages being studied were interviewed. The age group of the consultants ranges between 20-70 years old, and they are all males with various educational backgrounds and positions. While efforts were made to include female community leaders and potential participants, several female leaders were unavailable for interviews due to professional and personal commitments during the data collection period. Furthermore, cultural and social norms prevailing in the study area limited the researcher's ability to conduct one-on-one interviews with women. Thus, the interviewed sample of participants was all males. It should be noted that this limitation does not reflect any perceived difference in the level of linguistic knowledge between men and women, as women in the region possess comparable knowledge and experience regarding the local linguistic situation. Rather, the gender composition of the sample resulted from practical and socio-cultural constraints encountered during fieldwork. Recruitment was conducted through multiple channels:

- Community leaders and tribal elders: Tribal leaders (sheikhs and mukhtars) were contacted through established community networks and formal introductions facilitated by local civil society organizations.
- Educational professionals: Teachers and university professors with long-term residence in the district were identified through educational institutions and professional networks.
- Civil society activists: Individuals active in community organizations, cultural associations, and non-governmental organizations (NGOs) were recruited based on their documented knowledge of local demographics and language use.

Participants were selected based on the following criteria: (a) residence in the settlement for more than five years, (b) active engagement with diverse community members, and (c) demonstrated

knowledge of the linguistic composition of their settlement and neighbouring areas. To minimize bias, participants were not informed about responses from other interviewees in the same settlement, allowing for independent verification of data.

Table 1. The number of consultants interviewed for each language in Ninawa Province

	Language	Number of consultants
1	-Aramaic (Surith and Chaldean)	22
2	urdish (Northern and Central)	20
3	ni (Shabaki/Bajelani and Kakai)	20
4	Arabic (*qeltu* and *gelet*)	30

Data Collection Instruments

Data collection was conducted using a structured questionnaire adapted from Anonby, Taheri-Ardali, Haig et al.'s sociolinguistic study methodology (2020). The questionnaire included the following components:

- Settlement identification: Name of the settlement in Arabic and English, administrative classification (village, town, or city).
- Language inventory: List of all languages spoken in the settlement as mother tongues.
- Speaker proportions: Approximate percentage of the population speaking each language as a first language or primary home language.
- Multilingualism patterns: Information on secondary languages spoken by residents and contexts of language use.
- Demographic context: Approximate population size, presence of displaced populations, and recent demographic changes.

Here are the main research questions used to direct the components of the questionnaire mentioned above:

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

The questionnaire was administered through semi-structured interviews conducted in Arabic. Interviews lasted between 30 and 90 minutes. All interviews were conducted with informed consent, and ethical approval for the research was obtained from Carleton University Research Ethics Board (clearance number: CUREB-A 116207).

Settlement Selection and Coverage

Settlements were selected to achieve comprehensive coverage of al-Hamdaniyah District. The sampling frame was based on the Humanitarian Data Exchange (HDX) Iraq settlement database, which provides administrative classifications and geographic coordinates for settlements across Iraq. All settlements classified as cities, towns, or villages within al-Hamdaniyah District were included in the study, resulting in a total of 131 settlements surveyed. Fieldwork was conducted systematically by sub-district, with multiple visits to ensure access to participants despite security challenges and displacement. In cases where direct access to a settlement was not possible due to security concerns or ongoing conflict, interviews were

conducted with displaced residents or with individuals from neighbouring settlements who had recent knowledge of the linguistic composition.

Data Verification and Cross-Validation

To ensure reliability and accuracy, the researcher implemented a cross-validation protocol in which language data for each settlement were gathered from at least two independent informants whenever possible, and from three to five informants in larger towns and cities. Informants were not told what others from the same settlement reported, ensuring that responses were collected independently rather than shaped by prior answers. Where documentary evidence was available, reported patterns against census records, community reports, and published ethnographic studies were triangulated. Finally, each data point was assigned a confidence rating (high, medium, or low) based on the degree of agreement across informants and the presence of corroborating sources.

Data collection challenges included limited access to some settlements due to security conditions, the absence of residents in settlements affected by ISIS occupation (2014-2017), and the difficulty of estimating speaker proportions in settlements with high population turnover due to displacement. These limitations are noted in the discussion of results.

Geographic Data Collection and Integration

Geographic data, including latitudes, longitudes, and settlement names in English and Arabic, were collected from two primary sources:

- Humanitarian Data Exchange (HDX): The HDX Iraq settlement database (https://data.humdata.org/dataset/hotosm_irq_populated_places)
- provides open-access geographic information for settlements across Iraq, including administrative classifications, population estimates, and coordinate data. This database was used as the primary source for settlement locations and administrative hierarchies.
- Google Maps: For settlements not included in the HDX database or where coordinate data were missing, web scraping techniques were used to extract geographic coordinates from Google Maps. This involved querying settlement names and extracting latitude and longitude values from the returned results.

All geographic data were compiled into a structured Excel file with the following fields:

1. Unique ID: A hierarchical identifier created for each settlement using the format: IQ- [Province Code]- [District Code]- [Sub-district Code]- [Settlement Code]. For example, IQ-NI-HAM-BAG-001 represents settlement 001 in Baghdida sub-district, al-Hamdaniyah District, Ninawa Province.
2. Administrative hierarchy: Province (*muhāfadhah*), District (*qadā'*), Sub-district (*nāhīyah*), and Settlement (City/*madīnah* or Village/*qaryah*).
3. Settlement names: Arabic name, Romanized transliteration, and English name where applicable.
4. Coordinates: Latitude and longitude in decimal degrees (WGS84 coordinate system).
5. Language data: List of languages spoken in the settlement and approximate percentage of speakers for each language.

Data Integration and Analysis

The integration of fieldwork data with geographic data followed a sequence of analytical steps beginning with data preparation. Fieldwork responses were transcribed, and percentage estimates provided by multiple informants were averaged. Additionally, discrepancies were addressed through follow-up interviews when feasible or, where uncertainty remained, by assigning lower confidence ratings.

Next, a spatial database was constructed to enable GIS-based analysis. The Excel file containing settlement-level language data and geographic coordinates was imported into QGIS as a delimited text layer. Coordinate data were initially handled in the WGS84 geographic coordinate

system (EPSG:4326). For cartographic visualization, the language spoken by the largest proportion of residents in each settlement was designated as the dominant language, following established practices in language mapping (Anonby et al., 2019), while acknowledging that many settlements are multilingual.

Spatial analysis and map production were then carried out within QGIS (QGIS Development Team, 2024). Voronoi polygons were generated to estimate the spatial extent of language regions from settlement point data, partitioning space so that each polygon represents the area closest to a given settlement point (Aurenhammer, 1991; Okabe et al., 2000). This method offers a clear visual approximation while assuming Euclidean proximity and not accounting for population density, physical barriers, or within-settlement multilingualism (Thiessen, 1911; Dent et al., 2009). Multiple map types were produced, including point maps showing dominant languages by settlement, Voronoi polygon maps illustrating estimated language regions, and proportional maps displaying speaker percentages for individual languages across settlements. Base layers, including administrative boundaries and major cities, were sourced from the HDX Iraq administrative boundary shapefiles (Humanitarian Data Exchange, 2021), and all maps were exported as high-resolution images for publication.

Methodological Limitations

There have been several limitations that should be noted here. First, the estimation of speaker proportions relies on informed approximations provided by community members rather than comprehensive household surveys. Although cross-validation through multiple informants strengthens reliability, the resulting figures should be understood as best-informed estimates rather than census-level measurements.

Second, the dataset is temporally specific to the fieldwork period (2021–2024). Given the large-scale displacement and subsequent return movements following the ISIS conflict, language distributions in some settlements may remain in flux, and the patterns documented here may change over time.

Third, multilingualism is only partially captured in the cartographic outputs. While multilingualism data were collected, the maps prioritize dominant languages for each settlement to maintain visual clarity, which necessarily reduces the visibility of smaller speaker groups and overlapping language practices. In addition, security-related access constraints limited direct field access to some settlements. In these cases, information was obtained from displaced residents or knowledgeable individuals from nearby areas, which may introduce additional uncertainty.

Despite these limitations, the methodology offers a systematic and verifiable approach to documenting language distribution in al-Hamdaniyah District, and it represents a substantial improvement over earlier maps that relied primarily on secondary sources or generalized impressions.

5. A Reflection on Language Distribution

This section presents the core findings of the study, examining, analyzing, and describing patterns of language distribution and classification in al-Hamdaniyah District. It includes a detailed description of each language variety together with its classification. Figure 4 provides an overview of language distribution based on a map created using QGIS. This figure synthesizes the preparatory research described above by visualizing all language-distribution data collected for the study. Owing to the representational limits of static maps, the overview depicts only the language with the largest proportion of speakers in each settlement. More detailed figures follow, showing the estimated proportions of each language within individual settlements.

The linguistic landscape of the district is dynamic, shaped by ongoing displacement and return, shifting settlement patterns, and evolving social relations among communities. Such dynamism presents challenges for language-mapping: time-stamped observations can lag on-the-ground

change, and static polygon or point symbols may under-represent multilingualism and minority presences. Moreover, boundaries between speech communities are often porous rather than discrete. To reflect these realities in future mapping, the researcher plans to complement dominant-language depictions with proportional, settlement-level symbology to time-stamp observations and record sources and confidence ratings, to release versioned updates as new data are verified, and to pair quantitative layers with qualitative notes that document locally salient movements and narratives.

Settlements in al-Hamdaniyah District are not uniform as they vary in population size, internal heterogeneity, and degree of connectivity (e.g., proximity to larger urban centres and major roads versus relative isolation), as well as in other geographic and social features that influence present-day language distribution. These factors are relevant to interpreting the figures that follow and are referred to where appropriate in the discussion of settlements.

Because this study relies heavily on first-hand fieldwork and interview-based methods done by the first author (H. Ghanim), his positionality also warrants consideration. As someone born and raised in Mosul, with a particular linguistic and educational background, his position can facilitate access to some groups and make initial contact with others more difficult. To mitigate potential effects, a standard questionnaire was used, and more than one person was interviewed within each group. Prior interviews with subsequent respondents were not disclosed, and reports across groups and settlements were triangulated. Thus, convergent claims were treated as hypotheses to be cross-checked rather than as confirmation in their own right. The first author, H. Ghanim, remains attentive to how his background may shape both data collection and interpretation, and he flags potential influences where relevant in the analysis that follows.

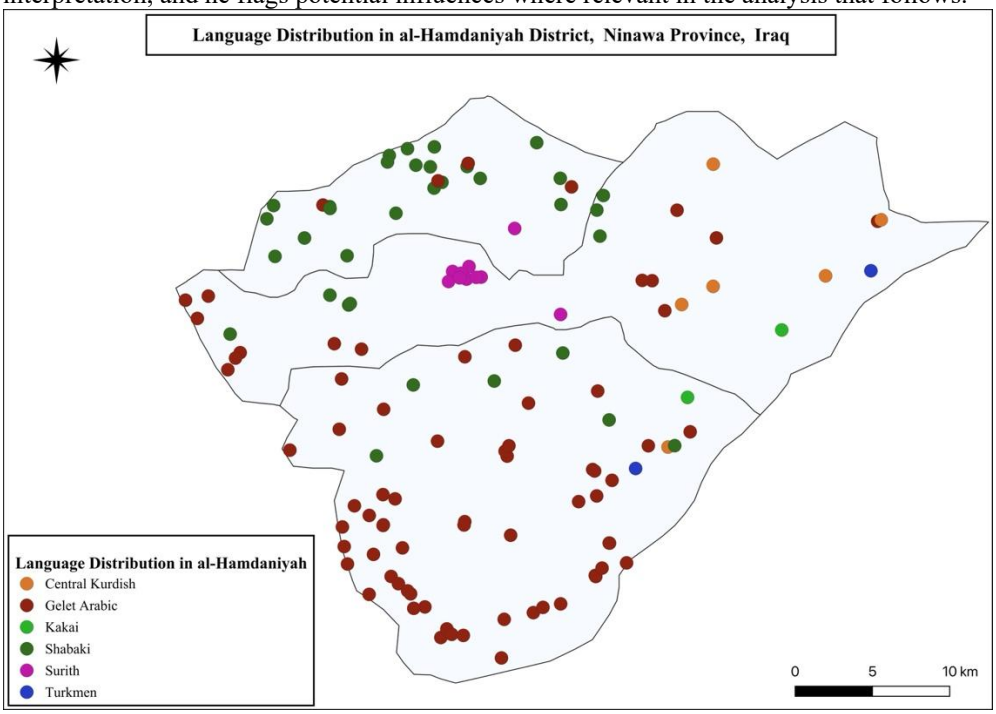

Figure 11. Overview of Language Distribution in al-Hamdaniyah District (Map by H. Ghanim, 2025. Data: HDX Iraq shape files and H. Ghanim's original fieldwork. Produced in QGIS)

Despite the district's small geographic extent, the figure reveals a striking multilingual mosaic.[1] The languages represented fall into three families, Semitic (*gelet* Arabic; Aramaic/Surith), Iranic (Central Kurdish [Sorani], Kakai, Shabaki), and Turkic (Turkmen), and a clear spatial pattern emerges. Arabic forms a broad outer belt encircling minority-language clusters, with concentrations along the borders with Mosul District and on the fringes abutting Kurdish-speaking districts such as Akre and al-Shekhan. Within this belt, Surith (Assyrian) is tightly clustered at the district centre, around Baghdida/Qaraqosh, Karamles, and Bartilla, long-standing centres of the district's Aramaic-speaking Christian communities. Shabaki appears as an inner ring around the Surith core and in additional east-central/south-eastern pockets, while Kakai occurs in adjacent clusters often close to Shabaki. Kurdish is concentrated near the northern and north-eastern boundaries, indicating border-proximal pockets rather than a broad interior presence. Turkmen appears only in isolated eastern/south-eastern settlements and is less widespread than the other groups. Interpreting these patterns requires attention to the cartography. The symbology plots the dominant mother tongue per settlement with equal-sized points, so it reflects counts of settlements rather than population size, suppresses bilingualism, and may leave interior blanks where fewer settlements are mapped rather than signalling true absence. In everyday practice, Arabic remains the language of schooling across the district. However, some church-affiliated schools offer classes in Assyrian, while Kurdish is taught in the Iraqi school system for two years in high school. In general, the map and analysis indicate an Arabic outer belt surrounding minority-language clusters, a concentrated Surith core, spatially structured Shabaki and Kakai zones, and sparse, boundary-proximal Kurdish and Turkmen presences.

Besides the point-based map presented in Figure 11 above, a Voronoi diagram (Voronoi Polygon map) is created to better visualize the linguistic situation in al-Hamdaniyah district (Figure 12).[2] It translates the point-based observations into continuous zones of "nearest-settlement influence." Each polygon is assigned the dominant mother tongue of the settlement at its seed point, so the map partitions the entire district into areas where, by straight-line proximity, that settlement (and thus its language) is closest. Compared with the point map, which counts settlements and leaves unfilled space between them, the Voronoi map makes three things much clearer: (i) the proximity and extent of language regions (e.g., a broad Arabic belt occupying much of the south and west; a compact Surith core centred on Baghdida/Qaraqosh, Karamles, and Bartilla. There are also structured Shabaki and Kakai blocks (small, border-proximal Central Kurdish and Turkmen zones); (ii) the edges where languages are most likely in contact, visible as the sharp polygon edges between Arabic–Shabaki, Shabaki–Kakai, and Surith–Arabic/Shabaki areas; and (iii) the areal share of each language (by surface area rather than by number of settlements), which the point map cannot convey. At the same time, the Voronoi approach imposes hard, straight boundaries and assumes Euclidean proximity. It does not incorporate population size, within-settlement multilingualism, physical barriers or roads, and it can inflate areas where settlements are sparse. The two maps are complementary as the point map shows where each settlement's dominant mother tongue is located, while the Voronoi map shows the spatial footprint and adjacency structure that those points imply. They highlight coherent cores (e.g., Surith), broad belts (Arabic), and likely contact zones among Shabaki, Kakai, Kurdish, and Turkmen communities.

[1] This map shows only languages spoken as a mother tongue, and only the main and most dominant languages in each area.

[2] Although many places are home to speakers of more than one language, only the language with the largest proportion of speaker is shown.

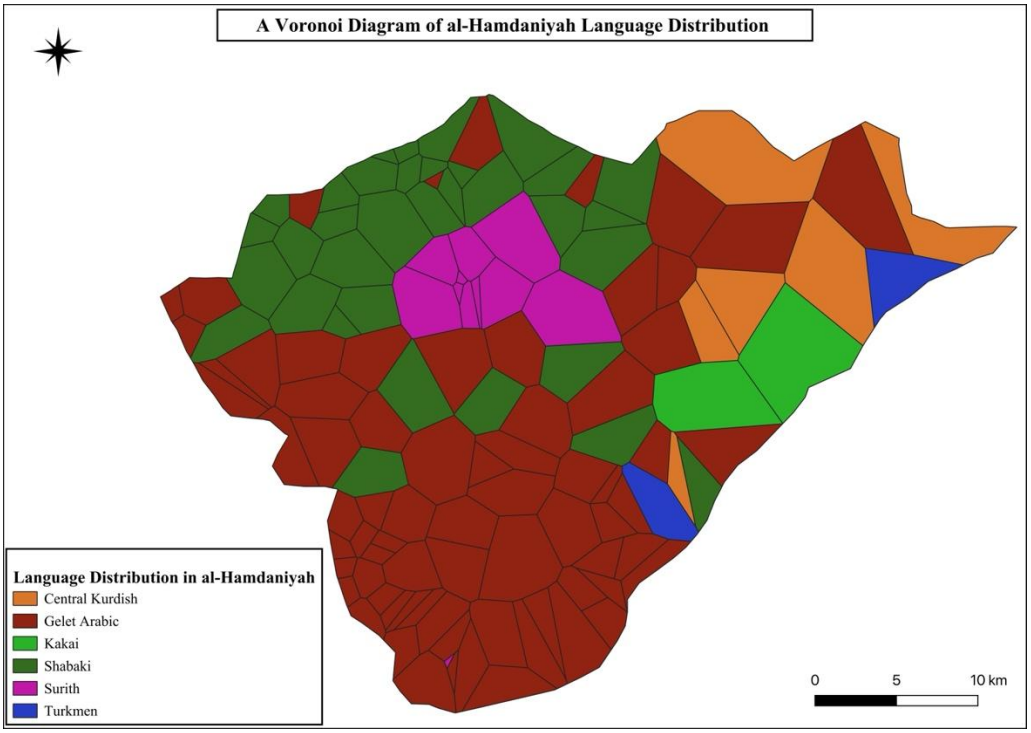

Figure 12. Voronoi polygon map of language distribution in al-Hamdaniyah District (Map by H. Ghanim, 2025. Data: HDX Iraq shape files and H. Ghanim's original fieldwork. Produced in QGIS)

Distribution and Classification of Individual Languages

Classifying languages is challenging, particularly when speakers use different varieties of the same language and identify those varieties with their place of residence or tribal affiliation. In al-Hamdaniyah, however, classification was easier because the languages spoken there are distinct and are not varieties of a single language. For instance, *gelet* Arabic is different from Assyrian, a Neo-Aramaic language. Although both belong to the Semitic language family, they constitute two different systems. The internal classification and distribution of each language found to be spoken in al-Hamdaniyah District is discussed below.

Aramaic

The map in Figure 13 shows the distribution of Aramaic in al-Hamdaniyah district with the percentages of the speakers. It is seen on the map that the number of speakers reduces as we move away from the center of al-Hamdaniyah district, known as Qaraqosh or Baghdida, where Christians lived for thousands of years. Again, the percentages of the speakers reflect the mother tongue speakers of the language and the languages spoken at home.

According to the language-distribution research on al-Hamdaniyah, Assyrian, a Neo-Aramaic language, is spoken in al-Hamdaniyah District. This variety derives from Aramaic, a Semitic language in the same family as Akkadian (Assyrian and Babylonian) (Millard, 2004). Assyrian, known as Syriac (the local name for Surith and this is the name used on the maps), is one of the varieties spoken in al-Hamdaniyah; it is a Semitic language from the Neo-Aramaic subgroup. The Neo-Aramaic varieties spoken in the north of Iraq are modern forms of Aramaic within the

Semitic family, and they belong to the North-Eastern Neo-Aramaic (NENA) subgroup of Neo-Aramaic. This subgroup is highly diverse, and it is spoken by Jews and Christians in Iraq, southeastern Turkey, and western Iran. Moreover, it has over 150 varieties, divided into Jewish NENA varieties and Christian NENA varieties. Surith is one of the Christian NENA varieties found to be spoken in several settlements in al-Hamdaniyah District (Khan, 2019). The NENA dialects do not directly descend from the earlier literary forms of Aramaic; rather, they are vernacular forms of Aramaic that were spoken in the region of northern Mesopotamia (Khan, 2007).

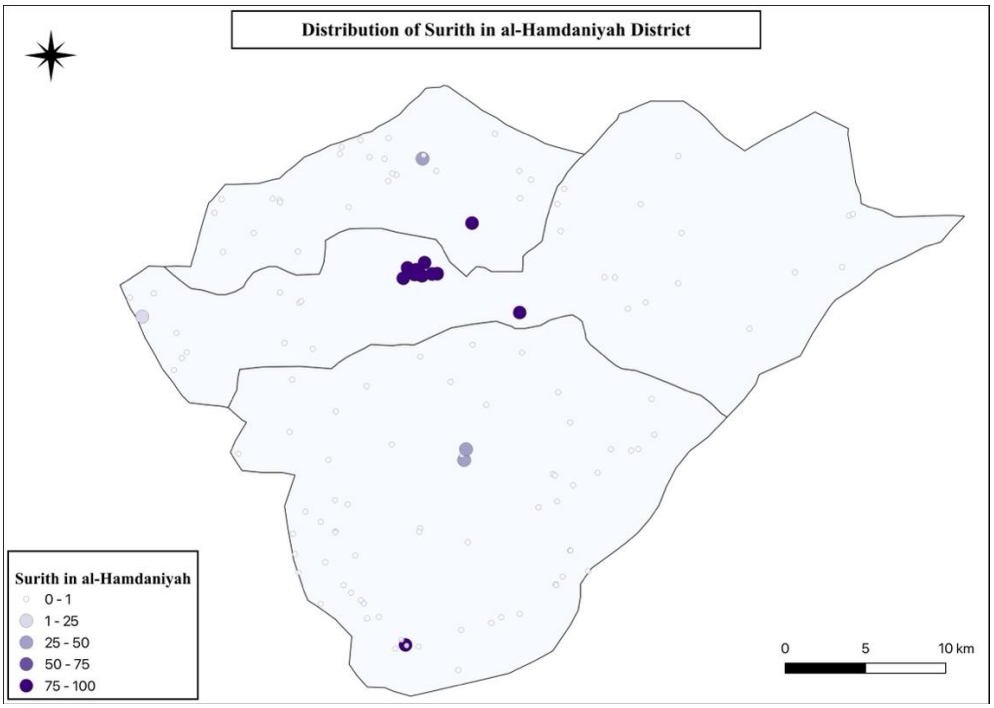

Figure 1. Percentages of Aramaic Speakers (Map by H. Ghanim, 2025. Data: HDX Iraq shape files and H. Ghanim's original fieldwork. Produced in QGIS)

Arabic

In al-Hamdaniyah district, the Arabic *gelet* variety is spoken, and the map below shows the distribution of this variety of Arabic in al-Hamdaniyah district. The map shows that Arabic is spoken by more speakers around the edges of the district than toward the center.

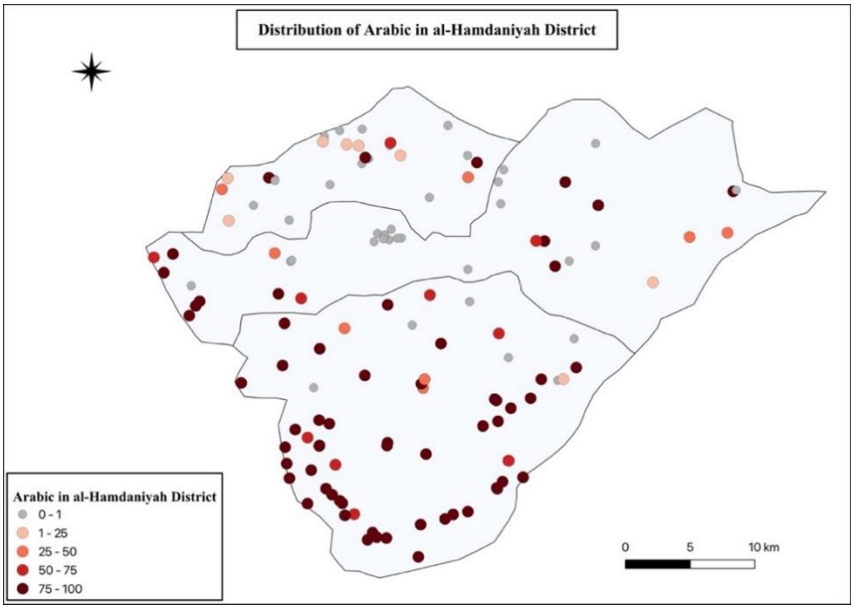

Figure 14. Percentages of the Gelet Arabic Variety (Map by H. Ghanim, 2025. Data: HDX Iraq shape files and H. Ghanim's original fieldwork. Produced in QGIS.)

Arabic is a Semitic language within the Afro-Asiatic language family. It has many modern varieties, and Iraqi Arabic is one of them. Moreover, most varieties spoken in northern Iraq belong to the so-called Tigris branch of the Mesopotamian group. In Iraq, two varieties of Arabic can be differentiated: an older sedentary-type set of dialects known as the *qeltu* varieties, and a more recent Bedouin-type variety known as the *gelet* variety. The *qeltu* varieties are mainly spoken in towns and ancient settlements, whereas the *gelet* variety is more recent (Prochazka, 2019).

Iranic Languages

Below is a map that shows the distribution of Northern Kurdish in several settlements in al-Hamdaniyah district. It is seen that Northern Kurdish is spoken in settlements on the border of Kurdistan Region and al-Hamdaniyah.

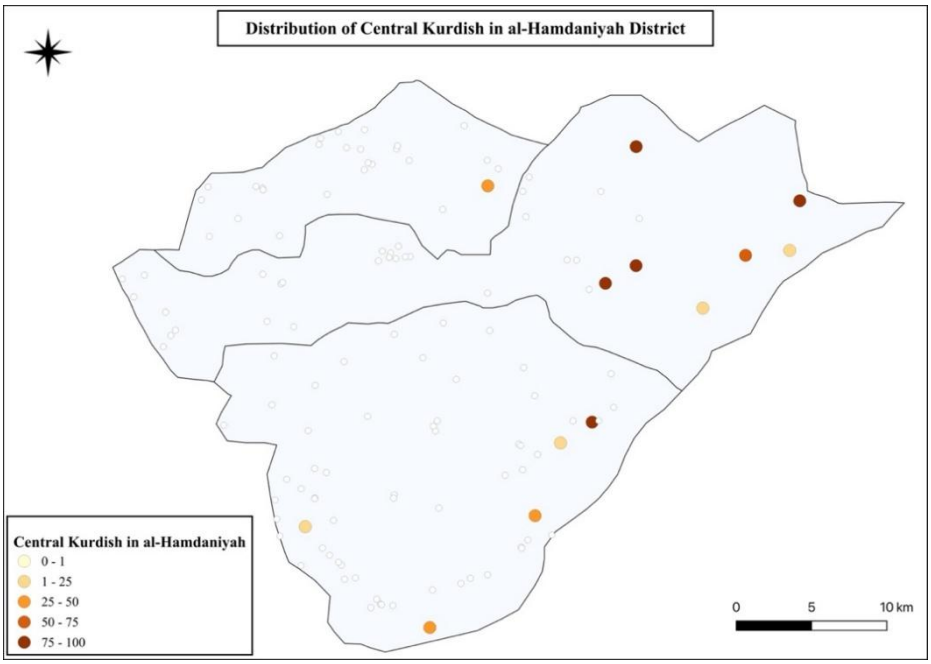

Figure 15. Distribution of Central Kurdish (Sorani) (Map by H. Ghanim, 2025. Data: HDX Iraq shape files and H. Ghanim's original fieldwork. Produced in QGIS.)

Central Kurdish: speakers of Kurdish call this dialects Sorani. Sorani is the official language used in official communications in Kurdistan Region (Allison, 2007). The Iranian languages spoken in northern Iraq fall into three main groups: Bahdini, the local name for the varieties of Northern Kurdish, also known as Kurmanji, Sorani Kurdish, which is the same as Central Kurdish, and Gorani. Bahdini, Sorani, and Gorani belong to the Northwestern branch of Iranian (Haig, 2019; Allison, 2007). Bahdini and Sorani are considered two varieties of Kurdish, and the relationship between the two is still not clear (Haig, 2019). Sorani is an official language in Iraq besides Arabic. Bahdini or Kurmanji, on the other hand, has less official status than Sorani. The Gorani varieties, on the other hand, have virtually no official status and must be considered highly endangered (Haig, 2019).

Based on fieldwork interviews, the term Kurmanji is also used instead of Northern Kurdish by its speakers to refer to the Kurdish spoken in Sinjar by the Ayzidis, who call it the Ayzidi language. As for Central Kurdish, it is known as Sorani or Hawleri by its speakers. However, Sorani is not known as Hawleri in Sulaymaniyah Province; it is simply referred to as Sorani.

Gorani: the Gorani group of Iranic languages that covers a cluster of varieties within the West Iranic languages spoken in Iraq and Iran, such as Hawrami, Kakai, and Shabaki/Bajelani. These varieties might have originated in the mountains of the Hawroman region of Western Iran (Mahmoudveysi, Bailey, Paul, and Haig, 2012). The other varieties, Shabaki/Bajelani, and Kakai, moved westward from the mountains of Hawroman to Ninawa Province, Iraq. The speech communities of these varieties are geographically scattered in the Iraq-Iran border region between Halabja and Khanaqin and extend northwest toward Mosul (Haig, 2019). Shabaki, whose speakers

are Shabaks, an ethno-religious minority group in Iraq who are socially and linguistically closely associated with the Bajelan. According to Mutar (2020), there are approximately 200,000 Shabakis in Iraq (750,000 according to the consultants that H. Ghanim met), mainly concentrated in the northern part of the country, specifically in Ninawa Plains. Bajelani, on the other hand, is associated with the speech of the Bajelan tribes and are spoken in villages near Khanaqin and Koy Sanjak to the southeast near the Iranian border (Haig, 2019). There are different views on Shabaki and Bajelani. Some consider them to be the same while others, like Bailey (2018), claim that there are differences between the two groups, and both groups do not identify themselves as the same in terms of language and religion (MacKenzie, 1956). As mentioned above, Haig (2019) mentions that Bajelani is a term that is normally related to the speech of the Bajelan tribes, and as a language, it is spoken in several locations east of Mosul and in villages near Khanaqin and Koy Sanjak to the southeast near the Iranian border. Haig (2019) mentions that the Bajelan are sometimes subsumed under the Shabak, but he adheres to MacKenzie (1956) in maintaining a difference between the two groups, which seems to be more in line with the self-perceptions of the speakers of these two Gorani varieties. Finally, Kakai, also known as Macho, is spoken by groups belonging to the Yaresan or *Ahl-e Haqq* religious community. Speakers of Kakai are in al-Hamdaniyah district in Ninawa Province as we all as around the towns of Topzawa near Kirkuk, Khanaqin and Arbil (2018).

Leezenberg (1994) states that the Gorani speakers he met in Iraq all considered themselves Kurds. This is, of course, not surprising since after centuries of close contact between speakers of Gorani and speakers of Central and Northern Kurdish. MacKenzie (1956) considers the phonology of Gorani basically identical to that of Sorani. There are good reasons for distinguishing Gorani from Central and Southern Kurdish, and most scholars of Iranian languages continue to do so by looking at verb paradigms as they are among the more reliable means of assessing historical relatedness. However, the consultants that the researcher met with have another view on Gorani varieties. Gorani or Hawrami, as known by its speakers, is used in the Hawraman region in Sulaymaniyah and in areas close to the Iraq-Iran border, and it has varieties like Shabaki/Bajelani and Kakai. Shabak live in the northeast of Mosul, with almost 750,000 people who speak Shabaki, which, according to them, is a non-official non-written language. They also add that it is a mix of Kurdish, Arabic, and Persian. The Shabaki speakers were divided on this matter; some of them said that it is not a Kurdish dialect and that it is an independent language, while others said it is a Kurdish dialect coming from Northern Kurdish (Kurmanji).

On the other hand, Bajelani is associated with the speech of the Bajelan tribes, and it is spoken in several locations east of Mosul, as well as in villages near Khanaqin and Koy Sanjak, to the southeast, near the Iranian border (2019). The Shabak consultants mentioned that Bajelani is the same as Shabaki as a language. However, speakers of Bajelani in Khanaqin and Zerbatia say that Bajelani is not the same language as Shabaki; rather, it is a language that comes from Kurdish. As for Kakais, speakers use the term Kakai to refer to their language.

Shabaki: It is spoken east of Mosul in the Ninawa plains area (Haig, 2019: 296). Shabaki is one of the most spoken languages in al-Hamdaniyah, as seen on the map below. It is spoken in the inner circle after Arabic *gelet*, and it encloses settlements where Aramaic is spoken in the center of the district.

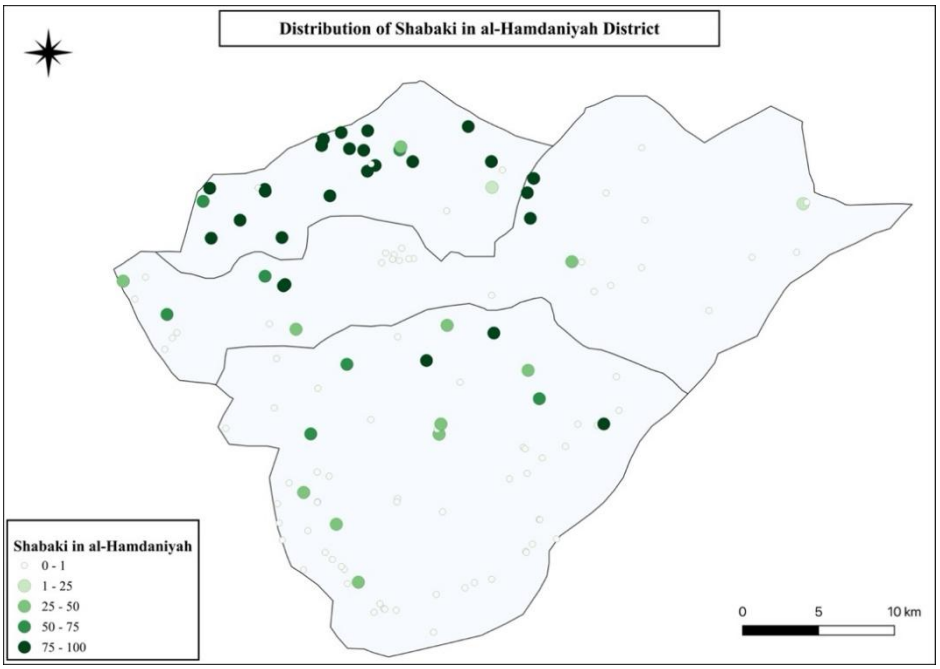

Figure 2. Distribution of Shabaki (Map by H. Ghanim, 2025. Data: HDX Iraq shape files and H. Ghanim's original fieldwork. Produced in QGIS)

Kakai: this is another Gorani variety spoken in Iraq. According to Haig (2019), the name Kakai is generally applied to groups belonging to the Yaresan religious community (Haig, 2019). Kakai is spoken in a couple of settlements in al-Hamdaniyah district, as shown on the map below.

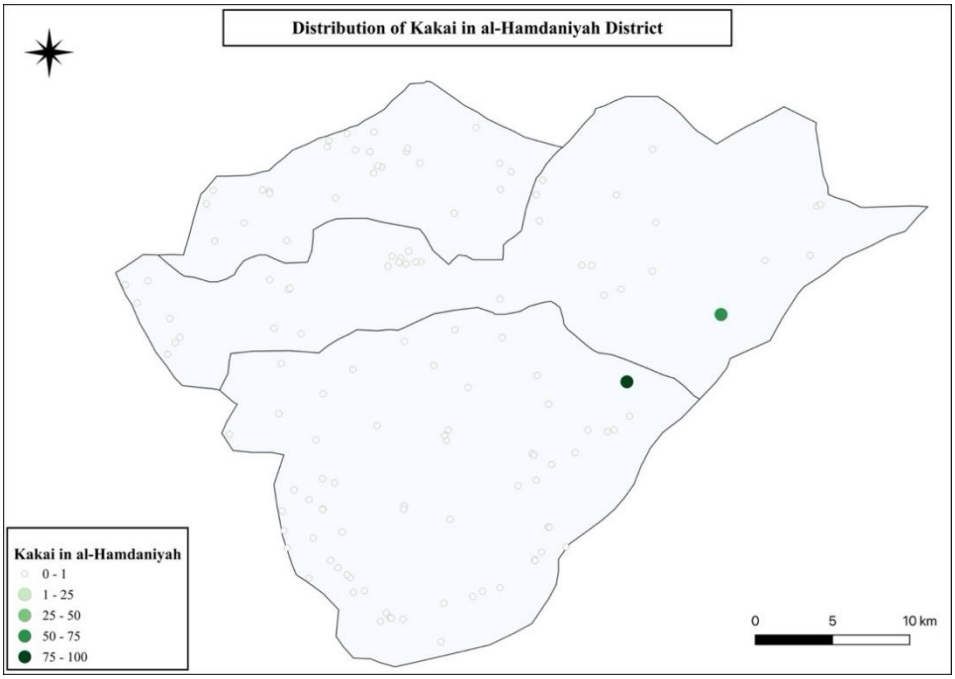

Figure 17. Distribution of Kakai (Map by H. Ghanim, 2025. Data: HDX Iraq shape files and H. Ghanim's original fieldwork. Produced in QGIS)

Turkic

There is only one variety of Turkic spoken in Iraq, known as Turkmen, spoken by the Turkmen of Iraq. Turkmen is spoken by the Turkmen who live in several separate areas, villages, and towns in Talafar, and they extend from the northwest to Bedre in the southeast. This means they are situated between Kurdish and Arabic-speaking areas (Bulut, 2007). Most Turkmen speakers in Iraq are bilingual or trilingual since Turkmen has no official status, even though the Turkmen form the third-largest group in Iraq after the Arabs and the Kurds.

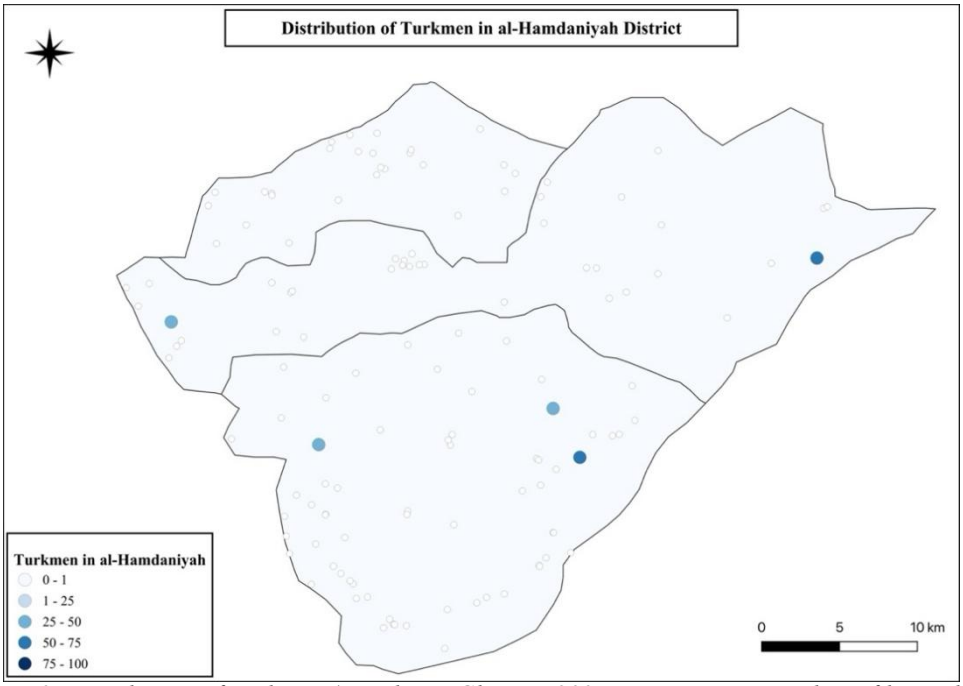

Figure 3. Distribution of Turkmen (Map by H. Ghanim, 2025. Data: HDX Iraq shape files and H. Ghanim's original fieldwork. Produced in QGIS)

6. Discussion

Spatial Patterns and Language Contact

The spatial distribution of languages in al-Hamdaniyah District reveals several patterns that have implications for understanding language contact, multilingualism, and language maintenance. The most striking pattern is the concentric arrangement of language communities, with Aramaic at the centre, Shabaki and Kakai forming an intermediate ring, and Arabic dominating the periphery. This pattern suggests historical processes of settlement and migration, with minority communities concentrated in central areas and the majority Arabic-speaking population surrounding them.

Moreover, the boundaries between language regions represent potential zones of language contact and multilingualism. The data collected in this study indicate that multilingualism is common in settlements near these boundaries. For example, in mixed settlements such as Bashiqa, some residents report fluency in multiple languages (Arabic, Kurdish, and Aramaic) due to daily interaction with speakers of different languages.

Furthermore, the ISIS occupation (2014-2017) and subsequent return of displaced populations have significantly affected language distribution. Some settlements that were previously majority Christian (Aramaic-speaking) have seen demographic changes, with lower return rates among Christians and settlement of Arab families in some areas. These changes may have long-term implications for language maintenance and shift.

Comparison with Previous Maps

The language distribution documented in this study differs in several respects from previous language maps of Iraq. One difference is that the maps in this study are more detailed. The settlement-level data collected in this study provide much greater spatial resolution than previous maps, which typically showed language distribution at the district or sub-district level. This fine-grained data reveal patterns of language distribution that are not visible at coarser scales.

Another difference is the time period of the maps. The data reflect the linguistic situation during the period 2021-2024, providing a time-stamped snapshot of language distribution. Previous maps often lacked clear temporal reference, making it difficult to assess changes over time. The time period covered by this study is particularly significant because it captures the post-ISIS period of return and reconstruction.

Finally, the methodological approach is different from the methodologies of the maps presented in the literature. Unlike previous maps that often relied on secondary sources or general knowledge, this study provides detailed documentation of data collection methods, sources, and limitations. This transparency allows for assessment of data quality and replication of methods in other areas.

Implications for Language Documentation and Policy

The findings of this study have several implications for language documentation, preservation, and policy. One important implication is language preservation. The data reveal that several minority languages in al-Hamdaniyah (Aramaic, Shabaki, Kakai) are spoken by relatively small populations in geographically concentrated areas. These languages face threats from displacement, emigration, and language shift. Documentation and preservation efforts should prioritize these languages, particularly given the ongoing out-migration of minority communities from Iraq.

Another implication is education and language rights. The multilingual character of al-Hamdaniyah District has implications for education policy. Minority language communities have advocated for mother-tongue education and for recognition of their languages in official contexts. The detailed documentation provided by this study can inform education planning and language policy by identifying where minority language speakers are concentrated and where mother-tongue education programs would be most beneficial.

Finally, Language distribution data can inform political representation and administrative arrangements. In Iraq's federal system, minority communities have sought recognition and representation in local government. Accurate documentation of where minority language communities are located can support these efforts.

Methodological Reflections

There are several methodological issues relevant to this study. A key challenge in this research was translating data from individual interviews into population-level estimates. The percentage estimates provided by informants represent their knowledge of language use in their communities rather than census-level data. To address this challenge, multiple informants were interviewed for each settlement when possible, and estimates were cross validated. However, it should be acknowledged that these are informed estimates rather than precise measurements.

Another challenge is the combination of HDX data and fieldwork. The methodology combined geographic data from the HDX Iraq settlement database with language data collected through fieldwork. The HDX data provided settlement locations and administrative classifications, while the fieldwork provided language data. The integration of these two data sources was straightforward for most settlements, but in some cases, there were discrepancies in settlement names or locations that required resolution through cross-referencing with Google Maps and consultation with local informants.

Additionally, there is a challenge relevant to the representation of multilingualism. The maps produced in this study primarily show the dominant language in each settlement. While this approach provides a clear visual representation of language distribution, it necessarily simplifies the complex multilingual reality of many settlements. Future research could explore alternative cartographic approaches that better represent multilingualism, such as pie charts at settlement locations showing the proportions of different languages, or multiple overlapping maps showing the distribution of each language separately.

Finally, there is the challenge of time, as language distribution is not static. The data collected in this study reflect the situation during the fieldwork period (2021-2024), but ongoing processes of displacement, return, emigration, and language shift mean that the linguistic landscape continues to change. Future research should include longitudinal studies that track changes in language distribution over time.

7. Conclusion

This study provides a detailed and comprehensive analysis of language distribution in al-Hamdaniyah District, Ninawa Province, Iraq. The research documents the spatial distribution of seven languages from three language families: Semitic (Arabic, Aramaic), Iranic (Northern Kurdish, Shabaki, Kakai), and Turkic (Turkmen) through systematic fieldwork and GIS mapping. In terms of speaker distribution across the 131 surveyed settlements, Arabic *gelet* is the dominant language with a presence in 68.5% of settlements and constituting the majority language in 56.2% of all settlements. Shabaki is the second most widespread language, spoken in 36.2% of settlements and dominant in 26.9% of them. Surith (Neo-Aramaic) is present in 12.3% of settlements and dominant in 9.2%. The remaining languages, Central Kurdish, Northern Kurdish, Turkmen, and Kakai, collectively account for the dominant language in approximately 7.6% of settlements, with Central Kurdish being the most prevalent among them 4.6%.

The findings reveal a complex multilingual mosaic characterized by a concentric spatial pattern, with Aramaic-speaking settlements concentrated at the center of the district, Shabaki and Kakai settlements forming an intermediate ring, and Arabic-speaking settlements dominating the periphery. This pattern reflects historical settlement dynamics, religious and ethnic identities, and the ongoing impacts of displacement and return following the ISIS occupation.

The methodology employed in this study, combining systematic fieldwork, cross-validation of data from multiple informants, integration of geographic data from open-access sources, and GIS mapping, provides a model for language documentation in other linguistically diverse and politically complex regions. The research emphasizes the importance of using proportional, settlement-level symbology, time-stamping observations, and recording sources and confidence ratings for future mapping efforts.

The study also recognizes that while static maps depict the dominant language in each settlement, the linguistic reality is complex, with multilingualism and porous boundaries between speech communities. Future research should explore methods for representing this complexity more fully, including longitudinal studies of language change, detailed ethnographic studies of multilingualism and language contact, and participatory mapping approaches that involve community members in the documentation process.

Al-Hamdaniyah District emerges from this study as the most linguistically diverse district in Ninawa Province, with Arabic forming a broad outer belt encircling minority-language clusters, a concentrated Aramaic-speaking core centred on Baghdida/Qaraqosh, Karamles, and Bartilla, and structured Shabaki and Kakai zones in between. This spatial mosaic reflects centuries of settlement history, religious and ethnic identity, and the demographic disruptions caused by the ISIS occupation of 2014–2017. By combining settlement-level fieldwork, cross-validated informant data, and GIS-based mapping, this study offers a significantly more granular and

methodologically transparent picture of language distribution in the region than previous maps provided. The findings carry direct implications for language preservation, mother-tongue education planning, and the political representation of minority communities. This research constitutes a foundational contribution to a broader project aimed at comprehensively mapping language distribution across all districts of Ninawa Province.

This research serves as a foundational step for a larger project aimed at mapping language distribution across all districts in Ninawa Province, contributing to a more accurate and nuanced understanding of the region's linguistic diversity. The findings have implications for language preservation, education policy, and political representation, providing data that can inform efforts to support minority language communities in Iraq.

Acknowledgement

I would like to express my sincere gratitude to the community members in al-Hamdaniyah who generously supported this research during my fieldwork. Their willingness to share their time, knowledge, experiences, and local insights was invaluable to the process of collecting data and understanding the linguistic and social landscape of the area. I am especially grateful for their hospitality, openness, and trust, which made this research possible. I also acknowledge the individuals who helped facilitate introductions, provided guidance in the field, and assisted with identifying relevant local information and settlement-level details. Their contributions greatly enriched the quality and accuracy of this study. Any errors or limitations in the interpretation and presentation of the data remain my own.

Statement on Conflict of Interest

The author declares that there are no conflicts of interest regarding the publication of this manuscript.

Appendix I – Fieldwork Interviews Questionnaire

Based on Anonby, Erik, Mortaza Taheri-Ardali, Geoffrey Haig, et al. "Atlas of the Languages of Iran (ALI) Questionnaire."[1]

General information

(to be filled out by the field researcher before or after fieldwork)

Name of person filling out this questionnaire: _____

Email address: _____

Location where the questionnaire was filled out: _____

Date filled out: _____

Name of person(s) checking / analyzing the questionnaire data (checker to fill this in): _____

Date checked / analyzed (checker to fill this in): _____

Date approved for upload (editors to fill this in): _____

Language variety (language, dialect, subdialect, etc.): _____

Language data from which province and which exact village(s)/town(s) _____

Is the data from your own fieldwork and/or from a published source? _____

If the data is from your own existing field notes, please ensure that it is from a single settlement. Then fill in the following details:

Field researcher(s):

Date of field research:

Location (province and district) of language under investigation:

Persian or English version of the questionnaire:

Language(s) of elicitation (indicate which are the main languages of elicitation, and which languages are used for clarification):

Sound recorder type _____ / Video recorder type _____ / Mic: internal or external?

If the data is from a published source, please fill in the following details:

Name of author(s):

Year of publication (if it's in Persian, put both the Iranian and Western year):

Title:

(if an article or book chapter) Page numbers:

(if an article or book chapter) Journal or Book Title:

(if in an edited volume) Editor(s):

Place of publication:

Publisher:

Any other important information:

During and after the interview, please use this page to write down any other information related to the questionnaire, for example:

- things to remember related to where audio and video files are stored
- observations on the quality of the recordings
- problems with the quality of the linguistic data
- observations on interesting linguistic structures that you encountered

[1] Anonby, Erik, Mortaza Taheri-Ardali, Geoffrey Haig, et al. "Atlas of the Languages of Iran (ALI) Questionnaire." Borealis, 2020. https://doi.org/10.5683/SP2/SDJ5N4

- tasks to follow up on
- etc.

Permission text and speaker details

My name is .... I come from .... I am a linguist from the University of .... I want to study and learn about the languages and dialects of [region] and help preserve knowledge about these languages for the future. I use the things that I learn for research and eventual publication in print, online in the Atlas of the Languages of Iraq, and in an online language archive. Do you give me the permission to ask you questions about your language, and to record and freely use what I learn?

[make sure the answer is loud and clear enough to be audible on the recording]

If the speakers have given their consent, say:

If you wish to discontinue this interview at any point or withdraw your consent for me to use anything that you've shared, including audio or video recordings, you are free to do so. I am able to withdraw your contributions until the point when they are published or archived.

Then, you can continue with the following questions, and the rest of the interview:

Would you like us to write your name(s) down, or would you like to remain anonymous?

_____

Name of speaker(s) (only put this if speakers have requested that their names be used):

(speaker 1) _____
(speaker 2) _____
(speaker 3) _____

Other details about the speakers (include this information for all speakers):

	Age:	Gender:	Profession:
(speaker 1)	_____	_____	_____
(speaker 2)	_____	_____	_____
(speaker 3)	_____	_____	_____

Languages spoken, along with:

a) speaker's L1 (first language learned from parents in the home), or L2+ (an additional language)?

b) speaker's stated degree of oral proficiency in each language (full = 5 / good = 4 / some = 3 / little = 2 / very little = 1)

c) frequency of usage in the past year (every day / at least once every week / at least once every month / at least once in the past year / never in the past year)

	Language:	L1 or L2+?	Proficiency:	Frequency:
(speaker 1)	_____	_____	_____	_____
	_____	_____	_____	_____
	_____	_____	_____	_____
	_____	_____	_____	_____
	_____	_____	_____	_____
(speaker 2)	_____	_____	_____	_____
	_____	_____	_____	_____
	_____	_____	_____	_____
	_____	_____	_____	_____
	_____	_____	_____	_____
(speaker 3)	_____	_____	_____	_____
	_____	_____	_____	_____

_____ _____ _____ _____
_____ _____ _____ _____
_____ _____ _____ _____

Languages the speaker can read and write, and the speaker's stated degree of written proficiency in each language (full = 5 / good = 4 / some = 3 / little = 2 / very little = 1)
(speaker 1) _____
(speaker 2) _____
(speaker 3) _____
Level of education (none / elementary / secondary / post-secondary):
(speaker 1) _____
(speaker 2) _____
(speaker 3) _____

Sociolinguistic information

The following portion of the questionnaire, which deals with language use in the context of a single settlement, is adapted from Anonby & Yousefian's (2011) sociolinguistic study.

Please answer the questions as you are able, providing estimates if necessary. You may mark any item for which you are unable to provide an answer, with "?".

8. What is the name of your community (city, town, village, etc.)? _____
9. What languages are spoken in the community as a mother tongue (that is, the first language a person learns at home as a child)? _____
10. What proportion (approximate percentage) of the population in the community speaks each of these languages as a mother tongue? _____
11. What do the people in the community call their own language(s) [the most general language grouping possible] and dialect(s) [more local groupings at the regional and local level, i.e., their particular variety of the language they mention] in their own language(s)? _____
12. What do the people in the community call their own ethnic group(s) in their own language(s)? [Ask the speaker to give ethnic group names at all relevant levels, from most general to most specific, as far as they are known, and specify each level in parentheses beside each of the ethnic names. _____
13. What is the name of the community (city, town, village, etc.) in each of the languages that are spoken there? _____
14. What proportion of the population (circle one answer for each):

speaks Arabic fluently.	all	most	some	a few	none
speaks some Arabic?	all	most	some	a few	none
speaks no Arabic?	all	most	some	a few	none

8. Besides their mother tongue(s) and Arabic, what additional languages and dialects do the people of the community speak? _____
9. What proportion (all/most/some/a few/none) of the children in the community are not learning the mother tongue of their parents as their mother tongue? _____
10. Which language communities do these children come from? (At the same time, specify the proportion of children – all/most/some/a few/none – who are not learning the mother tongue of their parents, for each language community.) _____
11. What language(s) are these children now learning as a mother tongue? _____

References

Allison, C. (2007). The Kurds are alive: Kurdish in Iraq. In J. N. Postgate (Ed.), *Languages of Iraq: Ancient and Modern* (pp. 135–158). British School of Archaeology in Iraq.

Anonby, E., Taheri-Ardali, M., Haig, G., et al. (2020). *Atlas of the languages of Iran (ALI) questionnaire* [Dataset]. Borealis. https://doi.org/10.5683/SP2/SDJ5N4

Aurenhammer, F. (1991). Voronoi diagrams — A survey of a fundamental geometric data structure. *ACM Computing Surveys*, *23*(3), 345–405. https://doi.org/10.1145/116873.116880

Bailey, D. (2018). *A grammar of Gawrajū Gūrānī* [Doctoral dissertation, Georg-August-University Göttingen]. https://doi.org/10.53846/goediss-7061

Blanc, H. (2024). *Communal dialects in Baghdad* (1st ed.). Brill.

Creason, S. (2004). Aramaic. In R. D. Woodard (Ed.), *The Cambridge Encyclopedia of the World's Ancient Languages* (pp. 391–426). Cambridge University Press. https://archive.org/details/cambridgeencyclo0000unse_t8o9/page/n5/mode/2up

Dent, B. D., Torguson, J. S., & Hodler, T. W. (2009). *Cartography: Thematic map design* (6th ed.). McGraw-Hill Higher Education.

Ghanim, H. (2026). Evaluation and analysis of Izady's map of the language distribution in Iraq. [Manuscript submitted for publication]

Ghanim, H. (2026). Mapping the languages of Ninawa Province in Iraq in 2025: An infographic. [Manuscript submitted for publication]

Google Maps. (2021). *Google Maps*. https://www.google.com/maps

Haig, G. (2018a). 2.3. Northern Kurdish (Kurmanjî). In G. Haig & G. Khan (Eds.), *The Languages and Linguistics of Western Asia* (pp. 106–158). De Gruyter. https://doi.org/10.1515/9783110421682-004

Haig, G. (2018b). 3.3. The Iranian languages of northern Iraq. In G. Haig & G. Khan (Eds.), *The Languages and Linguistics of Western Asia* (pp. 267–304). De Gruyter. https://doi.org/10.1515/9783110421682-009

Haig, G., & Khan, G. (Eds.). (2018). *The languages and linguistics of Western Asia: An areal perspective*. De Gruyter Mouton.

Haig, G., & Öpengin, E. (2014). Introduction to special issue — Kurdish: A critical research overview. *Kurdish Studies*, *2*(2), 99–122. https://kurdishstudies.net/menu-script/index.php/KS/article/view/59

Haig, G., & Öpengin, E. (Eds.). (2014). *Kurdish studies archive: Vol. 2, no. 2, 2014. Special Issue: Kurdish Linguistics*. Brill. https://doi.org/10.1163/9789004706552

Hassanpour, A. (1992). *Nationalism and language in Kurdistan, 1918–1985*. Mellen Research University Press.

Hiltermann, J. R. (2007). *A poisonous affair: America, Iraq, and the gassing of Halabja*. Cambridge University Press.

Human Rights Watch. (2003). *Iraq: Forcible expulsion of ethnic minorities* [Report]. https://www.hrw.org/report/2003/03/13/iraq-forcible-expulsion-ethnic-minorities

Humanitarian Data Exchange. (2021a). *Administrative boundaries* [Data set]. https://data.humdata.org/dataset/cod-ab-irq

Humanitarian Data Exchange. (2021b). *Iraq – Settlements* [Data set]. https://data.humdata.org/dataset/iraq-settlements

Izady, M. (2014). *Iraq: Linguistic composition in 2000* [Infographic]. School of International and Public Affairs, Columbia University. https://gulf2000.columbia.edu/images/maps/Iraq_Languages_lg.png

Jastrow, O. (2006). Arabic dialects in Turkey. In M. Eid, A. Elgibali, M. Woidich, & A. Zaborski (Eds.), *Encyclopedia of Arabic Language and Linguistics* (Vol. 1, pp. 117–123). Brill. https://dare.uva.nl/search?identifier=316bade0-d407-4b57-bae8-c4d1f1c899f6

Khan, G. (2007a). Aramaic in the medieval and modern periods. In J. N. Postgate (Ed.), *Languages of Iraq: Ancient and Modern* (pp. 95–114). British School of Archaeology in Iraq.
Khan, G. (2007b). The north-eastern Neo-Aramaic dialects. *Journal of Semitic Studies*, *52*(1), 1–20. https://doi.org/10.1093/jss/fgl034
Khan, G. (2016). *The Neo-Aramaic dialect of the Assyrian Christians of Urmi*. Brill.
Khan, G. (2018). 2.5. The Neo-Aramaic dialects of eastern Anatolia and northwestern Iran. In G. Haig & G. Khan (Eds.), *The Languages and Linguistics of Western Asia* (pp. 190–236). De Gruyter. https://doi.org/10.1515/9783110421682-006
Leezenberg, M. (1994). The Shabak and the Kakais: Dynamics of ethnicity in Iraqi Kurdistan. *ILLC Research Report and Technical Notes Series*, *94–07*(X), 1–19. https://eprints.illc.uva.nl/id/eprint/630/1/X-1994-07.text.pdf
MacKenzie, D. N. (1956). Bājalānī. *Bulletin of the School of Oriental and African Studies*, *18*(3), 418–435. https://doi.org/10.1017/S0041977X00087930
Mahmoudveysi, P. (2012). *The Gorani language of Gawrajŭ, a village of West Iran: Texts, grammar, and lexicon*. L. Reichert.
Mahmoudveysi, P., Bailey, D., Paul, L., & Haig, G. (2012). *The Gorani language of Gawraju, a village of West Iran: Texts, grammar, and lexicon*. Reichert Verlag. https://doi.org/10.29091/9783752005332
McCarus, E. N. (2009). Kurdish. In G. Windfuhr (Ed.), *The Iranian Languages* (pp. 587–633). Routledge. https://doi.org/10.4324/9780203641736
McDowall, D. (2021). *A modern history of the Kurds* (4th ed.). I.B. Tauris.
Mutar, S. (2020). *Mīsobotāmyā, mawsūʕat al-lughāt al-ʕirāqīya* [Mesopotamia, Iraqi Arabic encyclopedia]. Al-Ummah al-'Iraqiya Research Centre.
Natali, D. (2007). *The Kurds and the state: Evolving national identity in Iraq, Turkey, and Iran* (1st ed.). Syracuse University Press.
Okabe, A., Boots, B., Sugihara, K., Chiu, S. N., & Kendall, D. G. (Eds.). (2000). *Spatial tessellations*. John Wiley & Sons. https://doi.org/10.1002/9780470317013
Palva, H. (2009). *From qəltu to gələt: Diachronic notes on linguistic adaptation in Muslim Baghdad Arabic*. https://api.semanticscholar.org/CorpusID:190607438
Paul, L. (2008). Kurdish language I: History of the Kurdish language. In *Encyclopaedia Iranica* (pp. 1–22). https://www.iranicaonline.org/articles/kurdish-language/kurdish-language-i/
Postgate, J. N. (Ed.). (2007). *Languages of Iraq, ancient and modern*. British School of Archaeology in Iraq.
Postgate, J. N., & Bulut, C. (2007). Iraqi Turkmen. In J. N. Postgate (Ed.), *Languages of Iraq, Ancient and Modern* (pp. 159–187). British School of Archaeology in Iraq.
Procházka, S. (2018). 3.2. The Arabic dialects of northern Iraq. In G. Haig & G. Khan (Eds.), *The Languages and Linguistics of Western Asia* (pp. 243–266). De Gruyter. https://doi.org/10.1515/9783110421682-008
QGIS Development Team. (2024). *QGIS geographic information system* [Computer software]. Open Source Geospatial Foundation. https://qgis.org
Ruggles, S., Cleveland, L. L., Lovatón Dávila, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D. E., Heimann, Q., & Lee, J. (2024). *Integrated public use microdata series, international: Version 7.5* [Dataset]. IPUMS. https://doi.org/10.18128/D020.V7.5
Thiessen, A. H. (1911). Precipitation averages for large areas. *Monthly Weather Review*, *39*(7), 1082–1089. https://doi.org/10.1175/1520-0493(1911)39%3C1082b:PAFLA%3E2.0.CO;2
Woodard, R. D. (Ed.). (2004). *The Cambridge encyclopedia of the world's ancient languages*. Cambridge University Press.