Towards a Complete Mapping of Kurdish Dialectology

Christian Schuler[1], Raman Ahmad[2]

[1]Universität Hamburg [2]Hochschule für Angewandte Wissenschaften Hamburg

{christianschuler8989,raman.ahmad2022}@gmail.com

The Kurdish language and its many variations have been actively researched [1] with little consensus about their classifications, not in the least due to geographical, cultural, religious and political circumstances. One of many examples are Gorani and Zazaki, which can be considered to be standalone languages [2, 3], or part of Kurdish [4, 5], which might be one dialect [6, 4], or separate dialects [7, 8]. This work aims to shed light on Kurdish dialectology via creating an interactive map (see Figure 1).

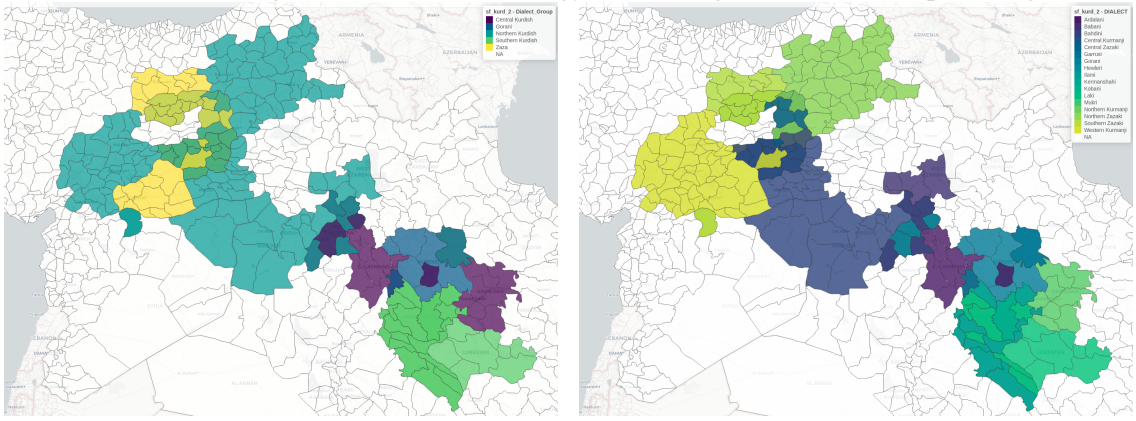


Figure 1: Examples from the current prototype. Left: Dialect groups layer. Right: Dialects layer.

Early work on Kurdish dialect geography [3], proposed a division of the Kurdish language into the three dialect groups Northern, Central and Southern Kurdish, which have been the basis for much of the linguistic work for Kurdish. Later sub-classifications took place Northern Kurdish (Kurmanji) first into five [9], later adjusted into three [10] distinct dialect groups. New understanding grows upon and rectifies prior work as can be seen by [11] writing about the often referenced work of [12]: *"it seems obvious that the seven subgroups identified by (Fattah, 2000) cannot be treated as clear-cut, sharply-defined bundles of dialects. They show considerable internal variation and in some cases seem to include dialects manifesting highly diverging features. Some varieties cannot be easily ascribed to any specific subgroup, but are transitional between subgroups."* The here allured concept of dialect continuum is crucial for understanding or properly displaying the Kurdish language on a map. Work on Kurdish language, regardless of producing a map or not, tends to focus on just a few dialects or one of the major dialect groups: Northern Kurdish [9], Central Kurdish [13], Southern Kurdish [11] and other dialects [14]. Be it personal interest, or borders between countries- for some reason most works fail to deliver more than a very rough outline of the Kurdish language landscape to interested readers.

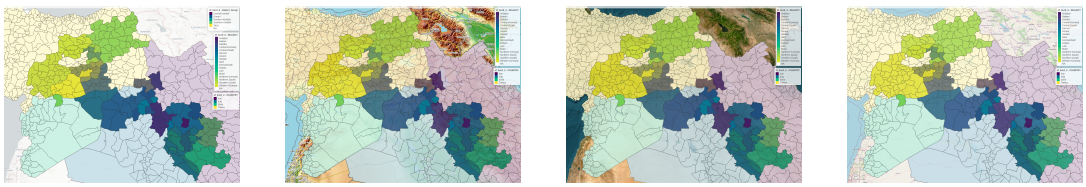


Figure 2: Examples from the current prototype, displaying different settings of the map.

At the time of writing, we have collected geographical information for 3551 distinct locations, considering the second layer of administrative boundaries of the countries Iran, Iraq, Syria and Turkey. Reviewing numerous related scientific publications rewarded us with a list of over 100 dialect names of varying granularity and specificity in regard of their definition, geographic position and linguistic classification.

1. Technical setup	Digital aspects Accessibility	**4. The data**	Data sources Data selection
2. Context and theme	Social context Map language(s)	**5. Visualisations & representational strategies**	Canvas Visualisation of key data Overlap of points/polygons
3. Type of language map	Language varieties treated Diachronic aspects	**6. Critique**	Possible motivation and bias Limitations

Table 1: Currently considered key aspects of the Evaluative Language Mapping Typology (ELM-T).

In a next step, we will integrate the Evaluative Language Mapping Typology (ELM-T) proposed by [15] for a more reflective map creation process. For the current focus of our work refer to Table 1.

Numerous aspects need to be considered, like using a colorblind-friendly palette [16] to provide high accessibility or how to manage the resolution of visual complexity given many overlapping dialects (see Figure 2) and diverging classifications. Collectible data is manifold but inconsistent, requiring a well considered storage structure. Finally, server solutions will be explored for online availability.

References

[1] S. Gündoğdu, E. Öpengîn, G. Haig, and E. Anonby, Eds., *Current Issues in Kurdish Linguistics*, ser. Bamberg Studies in Kurdish Linguistics (BSKL). University of Bamberg Press, 2019, vol. 1.

[2] V. Minorsky, "The Tribes of Western Iran," *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, vol. 75, no. 1/2, pp. 73–80, 1945.

[3] D. N. Mackenzie, "The Origins of Kurdish," *Transactions of the Philological Society*, vol. 60, no. 1, pp. 68–86, Nov. 1961.

[4] H. Salam Khalid, "Kurdish Language, its Family and Dialects," *International Journal of Kurdiname*, Apr. 2020.

[5] H. Hassani, "Can Linguistic Distance help Language Classification? Assessing Hawrami-Zaza and Kurmanji-Sorani," Oct. 2021.

[6] F. Hama-Khurshid, "Kurdish language and its dialects, a geographic analysis (Zmani Kurdi u diyalektakani, shikrdnawayaki jugrafyayi)," 2008.

[7] J. Sheyholislami, "Language Varieties of the Kurds," in *The Kurds: History, Religion, Language, Politics*. Vienna, Austria: Austrian Federal Ministry of the Interior, Jan. 2015, pp. 30–51.

[8] M. Azzat, K. Jacksi, and I. Ali, "The Kurdish Language Corpus: State of the Art," *Science Journal of University of Zakho*, vol. 11, no. 1, pp. 125–131, Feb. 2023.

[9] E. Öpengin and G. Haig, "Regional variation in Kurmanji : A preliminary classification of dialects," *Kurdish Studies*, vol. 2, Oct. 2014.

[10] G. Haig and E. Öpengin, "Kurmanji Kurdish in Turkey: Structure, varieties, and status," *Linguistic minorities in Turkey and Turkic-speaking minorities of the periphery*, pp. 157–229, 2018.

[11] S. Belelli, "Towards a dialectology of Southern Kurdish: Where to begin?" in *Current Issues in Kurdish Linguistics*, Dec. 2019, pp. 73–92.

[12] I. K. Fattah, *Livre : Les dialectes kurdes méridionaux, Acta iranica : encyclopédie permanente des études iraniennes, écrit par Ismaïl Kamandar Fattah - Peeters*. Peeters Publishers, 2000.

[13] S. Malmasi, "Subdialectal Differences in Sorani Kurdish," in *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3)*. Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016, pp. 89–96.

[14] S. Ahmadi, "Building a Corpus for the Zaza–Gorani Language Family," in *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*. Barcelona, Spain (Online): International Committee on Computational Linguistics (ICCL), Dec. 2020, pp. 70–78.

[15] A. Stone and E. Anonby, "A Typology for Evaluating Language Maps," in *Handbook of the Changing World Language Map*, S. D. Brunn and R. Kehrein, Eds. Cham: Springer International Publishing, 2022, pp. 1–30.

[16] Y. Ichihara, M. Okabe, K. Iga, Y. Tanaka, K. Musha, and K. Ito, "Color Universal Design-The Selection of Four Easily Distinguishable Colors for all Color Vision Types-," *Proceedings of SPIE - The International Society for Optical Engineering*, Jan. 2008.