Razi University

Research in Western Iranian Languages and Dialects

ISSN Print: 2345-2579 ISSN Online: 2676-573X

Linguistics Society of Iran

Toward a Phylogenetic Taxonomy of the Iranian Languages: A Focus on Kurdish Varieties and Other Related Dialects

Abdolrahman Sadeghi[1] and **Gholamhosein Karimi Doostan**[2]✉

1. Ph.D. Student, Department of Linguistics, Faculty of Literature and Humanities, Tehran University, Tehran, Iran. E-mail: Sadeghi68@ut.ac.ir
2. Corresponding Author, Professor, Department of Linguistics, Faculty of Literature and Humanities, Tehran University, Tehran, Iran. E-mail: gh5karimi@ut.ac.ir

Article Info

Article type:
Research Article

Article history
Received: 17 Sep 2024
Received in revised form: 30 Oct 2024
Accepted: 16 Nov 2024
Published online: 24 Sep 2025

Keywords:
phylogenetic taxonomy,
syntactic parameters,
syntactic distance,
PCM,
Iranian languages

ABSTRACT

Because of their objective measurement and unambiguous correspondences, syntactic parameters are used as comparanda for deep comparison and relatedness of language varieties. The Parametric Comparison Method (Longobardi & Guardiano, 2009) is based on the hypothesis that syntactic parameters are appropriate traits to use as comparanda for contrastive analysis and distance measurement of language pairs and then to provide language taxonomy. This article addresses the syntactic distance and genealogical taxonomy of language varieties that are closely related genealogically, geographically, and culturally. To analyze relatedness and genealogical issues of Kurdish language varieties and the Persian language, abstract grammatical rules, modeled as syntactic parameters which are applied in the domain of determiner phrase (DP), are used as comparanda. The results show that Laki and Lori are at a close distance from Sorani, and on the other hand, Kurmanji and Kırdki are also close syntactically. According to the tree achieved from the syntactic distances, Kurmanji is at one end of the spectrum and Persian is at the other end. Thus, it can be concluded that some preliminary deductions regarding the phylogenetic taxonomy of the Iranian languages and dialects, with a focus on Kurdish language varieties and other related dialects, can be proposed.

Cite this article: Sadeghi, A., & Karimi Doostan, G. (2025). Toward a phylogenetic taxonomy of the Iranian languages and dialects: A focus on Kurdish varieties and other related. *Research in Western Iranian Languages and Dialects*, *13*(3), 1–22. http//doi.org/ 10.22126/jlw.2024.11124.1797 (in English).

DOI: http//doi.org/ 10.22126/jlw.2024.11124.1797

Publisher: Razi University

1. Introduction

It is widely assumed that formal syntactic theories have barely played an important role in modern language classifications. However, most linguistic classifications and comparative analyses have been based on typological classification. The contribution of theories of formal syntax to linguistic phylogeny, with a series of works by Roberts (1993), Lightfoot (1999), and Longobardi (2001) among others, began to challenge the assumption. Longobardi and Guardiano (2009) suggested that the Parametric Comparison Method (PCM), which is based on syntax within the framework of modern generative theory, offers language distances and introductory taxonomies outside and inside of the Indo-European language family.

This article uses formal syntax to classify and determine syntactic distances of linguistic systems of some closely related Iranian languages and dialects. To achieve this goal, parameter values are used as *comparanda* (Roberts, 1998). These values are also used as taxonomic factors for historical reconstruction (Longobardi, 2003). More specifically, Longobardi and Guardiano's (2009) PCM approach has successfully been applied to the taxonomy of Indo-European languages (Longobardi et al., 2013) as well as other European languages (Longobardi et al., 2015a). In this article, the PCM approach, according to which syntax can encode the phylogenetic taxonomy of languages, is employed as a tool to yield syntactic distances and phylogenetic taxonomy of the Iranian languages, with a focus on different Kurdish varieties and the Persian language.

The term Kurdish is used to refer to a wider scope of language varieties in a specific geographical area. The Kurdish sprachbund or Kurdish dialects can refer to language varieties spoken in this area (Matras et al., 2022). Sprachbund is a linguistic convergence area that may be a more proper term to refer to language varieties addressed in this paper including Kurmanji, Sorani, Kalhori, Laki, Gorani, Kırdki, and Lori.[1] As the term sprachbund denotes, the linguistic area encompasses a geographically adjacent language bundle that may or may not have a genetic relationship but shows common features. Persian, as a language closely related to Kurdish, is also considered in this article.

The present article begins with a PCM presentation in section 2. After a brief description of the language varieties studied in this article, section 3 presents the methodology including the data gathering, samples, and the UPGMA algorithm used to measure the syntactic distances and to generate the language tree based on the quantitative results. Subsequently, some parameter manifestations are analyzed to determine the parameter values in each language variety in selected instances. Using these parameter values, the syntactic distances between pairs of language varieties are measured. Once the measurements are calculated, the phylogenetic classification is introduced in section 4. Finally, section 5 presents the conclusion of the article.

2. Theoretical Framework and Methodology

There are around 6,000 languages spoken in the world, and they differ greatly from one another (Baker, 2008). This results in wide linguistic diversity at different levels, including morphosyntactic, phonetic, semantic, and pragmatic differences. Researchers have classified

1. To avoid confusion or misunderstanding, by Kurmanji we mean Amed-Kurmanji, by Sorani we mean Suleymani-Sorani, by Kalhori we mean Kermashan-Kalhori, by Lori we mean Bakhtiyari-Lori, by Laki we mean Delfan-Laki, by Gorani we mean Pawah-Gorani (Hawrami), and by Kırdki we mean Sevirek-Kırdki (other terms: Zazaki, Dımıli, and Kırmanjki), and by Persian language we mean Tehrani-Farsi.

languages into clusters and families. These classifications have been mostly based on lexical or phonetic comparative studies (Hale, 2007). However, the morphosyntactic quantitative approaches to language classifications have barely been applied. The classification based on morphosyntactic features is interesting because these features are more robust to change than the phonetic or semantic features (De Gregorio et al., 2024). Despite superficial diversities among languages, they may be similar in some ways. Kurdish, for instance, is quite different from English or Arabic. Even though each structure of a Kurdish sentence is superficially different from its corresponding sentence in English, and each English sentence is distinct from its corresponding sentence in Arabic, the "formulas" of that sentence in the three mentioned languages may underlyingly differ in just one factor. Such factors combine and interact in different ways, resulting in linguistic diversity that can be seen cross-linguistically. Linguists have discovered these underlying factors or parameters. The role of parameters in linguistics is similar to the role of genes in molecular biology and that of atoms in chemistry (Baker, 2008; Longobardi & Guardiano, 2009).

Since the nineteenth century, historical linguistics and evolutionary biology have followed parallel paths, trying to classify languages and human populations, respectively. Both disciplines have tried to classify them into genealogically important families to compare the structures explaining the distribution of their similarities and to reconstruct their diverse origins (Croft, 2008; Orr, 1999).

In the case of the human population, the most classic way of classifying species is based on evidence that is externally accessible and observable, which is called morphological characters, such as the size and shape of the skull and body, the color of the eyes, hair, skin, and other traits. Since these features, due to evolutionary and natural selection from the environment, are unstable and change over time, they do not serve as entirely reliable taxonomic characters.

Due to theoretical advancements in biology, particularly the emergence of molecular genetics, phylogenetic studies have undergone a revolution in recent decades. The recently accessible molecular evidence has the major advantage of being less influenced by the changes caused by natural selection. Therefore, it is more likely to preserve genealogical information (Longobardi & Guardiano, 2009).

Like the classic way of classification in biology, the traditional investigation of phylogenetic relationships among languages in linguistics has focused on the most easily observable and accessible elements (in this case, whether inflections, affixes, or roots), which are sets of morphemes and words. These entities are termed *lexical* in a broad sense. For this reason, lexical items sharing similarities in both form and meaning appear to be able to provide the most convincing evidence for relatedness (Dunn, 2015). Linguistic classification has barely relied on the comparison of elements such as grammatical principles and syntactic rules, in particular.

Fundamentally, linguists have proposed two methods to identify genealogical relatedness, both of which are based on lexical comparison (Longobardi & Guardiano, 2009). These two methods are classical comparative methods and Greenberg's (1969) multilateral (or mass) comparison.

Since the classical comparative method relies on a small number of improbable phenomena, such as agreements in morphology and recurrent sound correspondences, it can lead to clear conclusions regarding language relatedness, with no need to resort to sophisticated measurement. In this manner, the mentioned method has mainly solved the issue

of establishing comparanda (Roberts, 1998) that are immune from chance similarities, without any detailed mathematical evaluation.

Despite its advantages, the classical comparative method is limited by a demarcation criterion. The method narrows the scope of inquiry to chronological spans and language sets. In such spans and sets, improbable phenomena such as recurrent correspondence are identifiable. It has provided a relatedness posterior verification of language families whose cognitive states were easily guessed before applying the method systematically. It is worth mentioning that this classical method has not been useful either as a heuristic or as proof for the long-distance classification of language families into deeper classes. It has also failed to be effective in recognizing lower taxa, that is, articulation within a language family.

The next notable effort to overcome the demarcation criterion limit is Joseph H. Greenberg's (1987, 2000) multilateral (or mass) comparison, which is very controversial. Even though this method does not depend on the exact sound correspondence criterion to recognize cognate sets, it still relies on lexical data. Greenberg suggests that the comparison of word lists across a large number of language sets simultaneously can compensate for the lack of exact sound correspondences. Greenberg has also claimed that precise diachronic steps and protolanguage reconstructions are not obligatory prerequisites to the phylogenetic taxonomy hypothesis, and synchronic difference/similarity consideration should be sufficient. Given an adequately universal meaning list, the method's advantage is that it can be applied to any language set, regardless of distance. However, the crucial disadvantage of Greenberg's method is that it fails to provide any exact measure of difference and similarity in meaning and sound.

So, it can be concluded that the classical comparative method and mass comparison method have serious restrictions. They are unable to give cogent proof for absolute historical relatedness and give precise measurements of classification distances. There is also a restriction of applicability. Thus, it is reasonable to claim that there has been barely a main accepted advancement in comparative methods based on the lexicon. It is not irrational to start to look at linguistic domains other than lexicon.

As in biology, theoretical developments in linguistics have been made with the emergence of formal and typological approaches to syntax (Chomsky, 1955, 1957a; Greenberg, 1963). The principles and parameters theory (Chomsky, 1981b) within the cognitive science framework has attempted to bring together the insights from both approaches regarding variation and grammatical universals. So, theoretical syntax, which studies the mind as a computation system of abstract symbolic entities, reached a new level of evidence suitable for classification and comparison.

Regarding these issues, syntax has never been a focal point in identifying the genealogical relationship among languages. For example, the relationships among Indo-European varieties have rarely been supported prevailingly or exclusively by syntactic evidence, since the identification of exact syntactic comparanda is difficult and syntax tends to be more invariable than the lexicon. This makes similarities less probative.

It is suggested that, in agreement with Roberts (1998), the principles and parameters theory provides the necessary systematic comparanda, that is, parameter values. Parametric methods for the diversity of grammar range from Chomsky's (1981b) principles and parameters to a minimalist approach to grammar (Longobardi & Treves, 2023). In the principles and parameters theory, parameters are understood as an open choice set between assumed binary values. Our invariant language faculty, Universal Grammar (UG), predefines these open

binary values, and each language learner based on the linguistic input from their environment determines the values. Open parameters define the variation space of biologically human grammars, and closed parameters specify each grammar. Thus, the acquisition of grammar can be reduced to parameter setting: every natural language grammar can be represented by a binary symbol string (0/1 or +/-). Each of these symbols encodes the parameter value of UG. These strings can easily be compared and used to establish exact sets of correspondences.

Identification of exact syntactic comparanda is achieved through a (morpho-)syntactic comparison method, which is a lexically blind system. Longobardi (2003) and Guardiano and Longobardi (2005) were the first to suggest the possibility of this kind of comparison, namely the Parametric Comparison Method (PCM).

The most important potential advantages of PCM include:

1. **formal properties**: Parameters form a universal list of distinct options. Because of this universal list, PCM requires only comparing the parameter values of the compared languages, leaving no doubt about which values are to be collated with each other.
2. **measuring syntax and lexicon**: PCM can provide precise measurements of distance between two language varieties, as well as mathematically grounded classifications.
3. **substantive properties**: Parameters are not affected by natural selection, that is, environmental factors. All community speakers seem to acquire the parameter values rather uniformly and unconsciously. So, intentional individual changes mostly fail to affect them.
4. **the implicational structure of linguistic diversity**: First, for a parameter to be valid, the grammatical traits followed from it all should be typologically co-vary. On the contrary, it is persuasive only if no other grammatical property co-varies with them. Second, as discussed in Baker (2008), a specific parameter value—but not the other—may imply the unrelatedness of another parameter.

An important step toward balanced results is choosing the parameters that are going to be collated. To choose reliable parameters, two factors should be considered: *exhaustiveness* and *practicality*. Exhaustiveness here means taking all possible parameters into account, while practicality refers to the applicability of these parameters. To investigate the properties of all possible parameters in a restricted subdomain, selecting an appropriate syntactic module will satisfy this requirement.

To achieve this goal, the considered module requires an exhaustive parameter set, a UG principle set that defines the implicational relationships of parameters and their scope, a trigger set for parameters values, and finally an algorithm to set parameters. Therefore, the determiner phrase (DP) internal syntax is chosen as the module to meet those conditions. The DP also has the additional advantage of being less complex in information structure in comparison with the clausal domain.

The most significant parameters that are applicable and meaningful to the Kurdish language varieties and Persian are chosen in this study. Since the language varieties are closely related in many aspects, studying and comparing all parameters may not provide valid results. Therefore, to achieve fine-grained results, some parameters are selected and applied to the varieties in detail. It may be necessary to elaborate on the surface structures of some parameters. A total of 49 meaningful parameter manifestations within the domain of DP have been identified. These parameters are listed in the first column of Figure 1. Some parameters of Figure 1 reflect current assumptions in typological or generative literature, with slight changes in their formulation.

Following Crisma et al. (2020), the Kurdish varieties and the Persian language are investigated within the PCM theoretical framework. To analyze the syntactic distances among the language varieties studied in this article, a set of 10 parameters in the domain of the determiner phrase (DP) that can be distinctive for the classification of languages is selected from Crismaet al. (2020, Appendix). A subset of these parameters, including the manifestations of the parameters, is analyzed to provide more precise results regarding the close language varieties considered in this study.

Data on the respective language varieties were mainly gathered from native speakers via interviews. A questionnaire, using English as a metalanguage, was employed to collect data so that potential trigger sets for each parameter could be identified. The translation of each English expression into the respective language varieties should allow the determination of its parameter value. If the parameter trigger in the questionnaire has two different structural representations in two language varieties, the values would be set in opposite ways.

The column in Figure 1 represents parameter values in each language variety as a vertical string showing parameter values. Parameter values are encoded as "+" or "-", indicating oppositional values only. Once they were collected and changed to the parametric values, they were analyzed in numerical terms. The first step involves calculating the number of differences and identities in the settings of parameters for each language pair. Then, these calculations are expressed as coefficients, pairs of positive integers represented as <i; d>. Here "d" stands for differences and "i" stands for identities. Eventually, for the purpose of representing a uniform distance ranking, the coefficients need to be reduced to a single numerical value. Hence, a coefficient reduction into a single numerical value was adopted. The relative distances of the language varieties are represented in Figure 2. The distances were obtained using the Jaccard-Tanimoto distance, in which the distance is calculated by dividing the number of differences by the sum of differences and identities (considering only positive identities), as follows:

$$\Delta \text{ Jaccard }_{(A, B)} = [N_{-+} + N_{+-}] / [N_{-+} + N_{+-} + N_{++}]$$

$$\text{Distance }_{(A, B)} = \frac{d}{d + i}$$

Therefore, two identical languages have a distance of 0, whereas two languages with compeletly opposite settings have a distance of 1. The other cases will fall in between. Such distances are indicated in Figure 2.

Once the relative distances are obtained, the syntactic tree for the relationships and distances of different Kurdish varieties—including Kurmanji, Sorani, Kalhori, Gorani, Kırdki (Zazaki)[1], Laki, and Lori[2]—and the Persian language is presented using the UPGMA algorithm.

The UPGMA algorithm (Sokal & Michener, 1958), which deals with distances, is used to produce ultra-metric data from syntactic distances, that is, trees constructed in accordance with the molecular clock hypothesis (Bromham & Penny, 2003; Felsenstein, 2004). It first clusters together the two varieties that exhibit the smallest distance, creates a binary-branching tree, and constructs a new node for the two language varieties clustered. Using the average distance between the two language varieties and the other varieties, it then computes

1. Zazaki has a pejorative connotation. Another term is Dim(i)li (Windfuhr, 2009).
2. There are still controversial debates on Kurdish language varieties. It is expected that considering Lori as a variety closer to Kurdish dialects rather than Persian dialect may raise questions. However, it should be noted that by Lori we mean Bakhtiari Lori and the outcomes of this study are merely based on DP data.

the new distance between this new node and the existing node, and this procedure is repeated until only one node remains. The length of each branch in the rooted tree, which is the distance between every single leaf and the root, is uniform. This method considers all units of the taxonomy as leaves, meaning it does not locate any language on ancestral (non-terminal) nodes (Longobardi et al., 2013).

3. Data Analysis and Discussion

A total of 49 manifestations of parameters employed for the recent wide-range taxonomy in the framework of PCM (Ceolin et al., 2020) are used to measure the syntactic distances among language varieties in the present paper (Figure 1). Figure 1 lists the parameters and parameter values in each variety.

For concreteness, the grammaticalized agreement, semantic person, and semantic gender parameters are presented in sections 1, 2, and 3, respectively. The grammaticalized agreement differentiates languages in which words agree in φ-features (gender, person, and number) from languages in which they don not. The semantic person parameter indicates person differences in pronouns, and the semantic gender parameter distinguishes between two third person pronouns that indicate perceived biological sex and/or animacy. To obtain finer-grained distinctions among the language varieties studied in this research, each parameter is divided into a subset where necessary. For instance, the grammaticalized agreement parameter can be studied in more detail. Every aspect of agreement within the determiner phrases is investigated to show distinctions among language varieties. These aspects include agreement on gender, person, and number between determiners—including demonstratives, (in)definite articles, quantifiers, and others—and head nouns. Some instances of these agreements are shown in the following sections. The values of each parameter in each language are then shown in Figure 1.

3.1 Grammaticalized Agreement

Grammaticalized agreement differentiates languages in which words agree in φ-features (e.g., Indo-European and Semitic languages) from languages that do not distinguish this difference (e.g., Mandarin and Japanese). To obtain finer-grained distinctions among language varieties, nominative and accusative case features are considered. Sections 3.1.1 and 3.1.2 present the agreements in number between demonstratives and head nouns in the nominative and accusative cases. Sections 3.1.3 and 3.1.4 present the agreements in gender between demonstratives and head nouns in the nominative and accusative cases. Finally, Sections 3.1.5 and 3.1.6 present the agreement in gender between adjectives and head nouns in the nominative and accusative cases.

3.1.1 Grammaticalized Agreement in Number Between Demonstratives and Head Nouns in the Nominative Case

This kind of agreement differentiates languages with regard to the number agreement between demonstratives and head nouns in the nominative case. If the distinction is established in a language variety, the parameter value for that variety will be positive. If it does not, the value will be negative. Consider the following examples:

Ex: 1. **This boy** sees me. ENGLISH
2. **These boys** see me.

1.	***ʔæv***	***kʊr***	*mɪn*	*dɪ-bin-æ*	KURMANJI
	this	boy	me	IND-see-3.SG	
2.	***ʔæv***	***kʊr***	*mɪn*	*dɪ-bin-ɪn*	
	this	boy	me	IND-see-3.PL	
1.	***ʔæm***	***kʊr-æ***	*mɪn*	*dæ-bin-e*	SORANI
	this	boy-SP	me	IND-see-3.SG	
2.	***ʔæm***	***kʊr-an-æ***	*mɪn*	*dæ-bin-ɪn*	
	this	boy-PL-SP	me	IND-see-3.PL	
1.	***ʔi***	***kʊr-æ***	*mɪ*	*dɪ-win-i*	KALHORI
	this	boy-SP	me	IND-see-3.SG	
2.	***ʔi***	***kʊr-ejl-æ***	*mɪ*	*dɪ-win-ɪn*	
	this	boy-PL-SP	me	IND-see-3.PL	
1.	***ʔi***	***kʊr-æ***	*ʔæmn-i*	*mæ-win-o*	GORANI
	this	boy-SP	me-OBL	IND-see-3.SG	
2.	***ʔi***	***kʊr-e***	*ʔæmn-i*	*mæ-win-an*	
	this	boy-PL	me-OBL	IND-see-3.PL	
1.	***no***	***laʒæk***	*mɪ*	*vin-æn-o*.	KIRDKI
	this	boy	me	see-IND-3.SG	
2.	***ne***	***laʒæk-i***	*mɪ*	*vin-æn-e.*	
	these	boy-PL	me	see-IND-3.PL	
1.	***ʔi:***	***kʊr-æ***	*mɪn-æ*	*mɪ-wn-i*	LAKI
	this	boy-SP	me-OBL	IND-see-3.SG	
2.	***ʔi:***	***kʊr-æl-æ***	*mɪn-æ*	*mɪ-wn-ɪn*	
	this	boy-Pl-SP	me-OBL	IND-see-3.PL	
1.	***ʔi:***	***kʊr-o***	*mo-ne*	*ʔi-bin-e.*	LORI
	this	boy-SP	me-OBL	IND-see-3.SG	
2.	***ʔi:***	***kʊr-gæl***	*mo-ne*	*ʔi-bin-en.*	
	this	boy-SP	me-OBL	IND-see-3.SG	
1.	***ʔin***	***pesær***	*mæn-o*	*mi-bin-e*	PERSIAN
	this	boy	me-OBL	IND-see-3.SG	
2.	***ʔin***	***pesær-a***	*mæn-o*	*mi-bin-æn*	
	this	boy-PL	me-OBL	IND-see-3.PL	

3.1.2 Grammaticalized Agreement in Number Between Demonstratives and Head Nouns in the Accusative Case

This kind of agreement differentiates languages with regard to the number agreement between demonstratives and head nouns in the accusative case. If the distinction is established in a language variety, the parameter value for that variety will be positive. If it does not, the value will be negative. Consider the following examples:

Ex: 1. I see **this boy**. ENGLISH
2. I see **these boys**.

1.	*ʔæz*	***vi***	***kʊr-i***	*dɪ-bin-ɪm*	KURMANJI
	I	this.OBL	boy-OBL	IND-see-1.SG	
2.	*ʔæz*	***van***	***kʊr-an***	*dɪ-bin-ɪm*	
	I	these.OBL	boy-PL	IND-see-1.SG	

1.	*mɪn*	***ʔæm***	***kʊr-æ***	*ʔæ-bin-ɪm*	SORANI
	I	this	boy-OBL	IND-see-1.SG	
2.	*mɪn*	***ʔæm***	***kʊr-an-æ***	*ʔæ-bin-ɪm*	
	I	this	boy-PL-SP	IND-see-1.SG	
1.	*mɪ*	***ʔi***	***kʊr-æ***	*dɪ-win-ɪm*	KALHORI
	I	this	boy-OBL	IND-see-1.SG	
2.	*mɪ*	***ʔi***	***kʊr-ejl-æ***	*dɪ-win-ɪm*	
	I	this	boy-PL-SP	IND-see-1.SG	
1.	*ʔæm*	***ʔi***	***kʊr-æ-jæ***	*mæ-win-u*	GORANI
	I	this	boy-OBL	IND-see-1.SG	
2.	*ʔæm*	***ʔi***	***kʊr-an-æ***	*mæ-win-u*	
	I	this	boy-PL-SP	IND-see-1.SG	
1.	*ʔæz*	***no***	***laʒæk-i***	*vin-æn-a.*	KIRDKI
	I	this.OBL	boy-OBL	see-IND-1.SG	
2.	*ʔæz*	***ne***	***laʒæk-an***	*vin-æn-a*	
	I	this.OBL	boy-PL	see-IND-1.SG	
1.	*mɪ*	***ʔi***	***kʊr-æ***	*mɪ-wn-ɪm.*	LAKI
	I	this	boy-OBL	IND-see-1.SG	
2.	*mɪ*	***ʔi***	***kʊr-æl-æ***	*mɪ-wn-ɪm.*	
	I	this	boy-PL-SP	IND-see-1.SG	
1.	*mo*	***ʔi***	***kʊr-ene***	*ʔi-bin-om.*	LORI
	I	this	boy-OBL	IND-see-1.SG	
2.	*mo*	***ʔi***	***kʊr-gæl-ene***	*ʔi-bin-om.*	
	I	this	boy-PL-SP	IND-see-1.SG	
1.	*mæn*	***ʔin***	***pesær-o***	*mi-bin-æm*	PERSIAN
	I	this	boy-SG-SP	IND-see-1.SG	
2.	*mæn*	***ʔin***	***pesær-a ro***	*mi-bin-æm*	
	I	this	boy-PL-SP	IND-see-1.SG	

As can be observed from the above data, only Kırdki represents grammaticalized agreement in number in both the nominative and accusative cases. Kurmanji represents this feature in the accusative case but doesn't represent it in the nominative case. This reveals one of the distinctions between these two language varieties. Grammaticalized agreement in number between demonstrative and head nouns is not represented among other varieties. Therefore, the parameter values for Kırdki will be positive in both cases and for Kurmanji will be positive in the accusative case but negative in the nominative case. In all other varieties, the parameter value will be negative.

3.1.3 Grammaticalized Agreement in Gender Between Demonstratives and Head Nouns in the Nominative Case

This type of agreement differentiates languages with regard to gender agreement between demonstratives and head nouns in the nominative case. If the distinction is established in a language variety, the parameter value for that variety will be positive. If it does not, the value will be negative. Consider the following examples:

Ex:	1. **This boy** sees me. 2. **This girl** sees me.				ENGLISH
1.	***ʔæv***	***kʊrɪk***	*mɪn*	*dɪ-bin-æ.*	KURMANJI
	this.NOM	boy	me.OBL	IND-see-3.SG	
2.	***ʔæv***	***kætʃɪk***	*mɪn*	*dɪ-bin-æ.*	
	this.NOM	girl	me.OBL	IND-see-3.SG	
1.	***ʔæm***	***kʊr-æ***	*mɪn*	*dæ-wen-e.*	SORANI
	this	boy-SP	me	IND-see-3.SG	
2.	***ʔæm***	***kæniʃk-æ***	*mɪn*	*dæ-wen-e.*	
	this	girl-SP	me	IND-see-3.SG	
1.	***ʔi***	***kʊr-æ***	*mɪ*	*dɪ-win-i.*	KALHORI
	this	boy-SP	me	IND-see-3.SG	
2.	***ʔi***	***dɪwæt-æ***	*mɪ*	*dɪ-win-i.*	
	this	girl-SP	me	IND-see-3.SG	
1.	***ʔi***	***kʊr-æ***	*ʔæmn-i*	*mæ-win-o*	GORANI
	this	boy-M.OBL	me-OBL	IND-see-3.SG	
2.	***ʔi***	***kɪnatʃ-e***	*ʔæmn-i*	*mæ-win-o*	
	this	girl-F.OBL	me-OBL	IND-see-3.SG	
1.	***no***	***laʒæk***	*mɪ*	*vin-æn-o.*	KIRDKI
	this.NOM	boy	me	see-IND-3.SG	
2.	***na***	***kæjnæk***	*mɪ*	*vin-æn-a.*	
	this.NOM	girl	me	see-IND-3.SG	
1.	***ʔi:***	***kʊr-æ***	*mɪn-æ*	*mɪ-wn-i.*	LAKI
	this	boy-SP	me-OBL	IND-see-3.SG	
2.	***ʔi:***	***dɪt-æ***	*mɪn-æ*	*mɪ-wn-i.*	
	this	girl-SP	me-OBL	IND-see-3.SG	
1.	***ʔi***	***kʊr***	*mo-ne*	*ʔi-bin-e.*	LORI
	this	boy	me-OBL	IND-see-3.SG	
2.	***ʔi***	***dowær***	*mo-ne*	*ʔi-bin-e.*	
	this	girl	me-OBL	IND-see-3.SG	
1.	***ʔin***	***pesr-e***	*mæn-o*	*mi-bin-e.*	PERSIAN
	this	boy-SP	me-SP	IND-see-3.SG	
2.	***ʔin***	***doxtær-e***	*mæn-o*	*mi-bin-e.*	
	this	girl-SP	me-SP	IND-see-3.SG	

3.1.4 Grammaticalized Agreement in Gender Between Demonstratives and Head Nouns in the Accusative Case

This type of agreement differentiates languages with regard to the gender agreement between demonstratives and head nouns in the accusative case. If the distinction is established in a language variety, the parameter value for that variety will be positive. If it does not, the value will be negative. Consider the following examples:

Ex: 1. I see **this boy**. ENGLISH
2. I see **this girl**.

1.	*ʔæz*	***vi***	***kʊr-i***	*dɪ-bin-ɪm.*	KURMANJI
	I	this.M.OBL	boy-M-OBL	IND-see-1.SG	
2.	*ʔæz*	***ve***	***kætʃɪk-e***	*dɪ-bin-ɪm.*	
	I	this.F.OBL	girl-F-OBL	IND-see-1.SG	
1.	*mɪn*	***ʔæm***	***kʊr-æ***	*ʔæ-bin-ɪm.*	SORANI
	I	this	boy-SP	IND-see-1.SG	
2.	*mɪn*	***ʔæm***	***kæniʃk-æ***	*ʔæ-bin-ɪm.*	
	I	this	girl-SP	IND-see-1.SG	
1.	*mɪ*	***ʔi***	***kʊr-æ***	*dɪ-win-ɪm.*	KALHORI
	I	this	boy-SP	IND-see-1.SG	
2.	*mɪ*	***ʔi***	***dɪwæt-æ***	*dɪ-win-ɪm.*	
	I	this	girl-SP	IND-see-1.SG	
1.	*ʔæm*	***ʔi:***	***kʊr-æ-jæ***	*mæ-win-u*	GORANI
	I	this	boy-SP-DEF	IND-see-1.SG	
2.	*ʔæm*	***ʔi:***	***kɪnatʃe***	*mæ-win-u*	
	I	this	girl.OBL	IND-see-1.SG	
1.	*ʔæz*	***ne***	***laʒæk-i***	*vin-æn-a.*	KIRDKI
	I	this.M.OBL	boy-M-OBL	see-IND-1.SG	
2.	*ʔæz*	***na***	***kæjnæk-e***	*vin-æn-a.*	
	I	this.F.OBL	girl-F-OBL	see-IND-1.SG	
1.	*mɪ*	***ʔi:***	***kʊr-æ***	*mɪ-wn-ɪm.*	LAKI
	I	this	boy-SP	IND-see-1.SG	
2.	*mɪ*	***ʔi:***	***dɪt-æ***	*mɪ-wn-ɪm.*	
	I	this	girl-SP	IND-see-1.SG	
1.	*mo*	***ʔi:***	***kʊr-ene***	*ʔi-bin-om.*	LORI
	I	this	boy-OBL	IND-see-1.SG	
2.	*mo*	***ʔi:***	***dowær-ene***	*ʔi-bin-om.*	
	I	this	girl-OBL	IND-see-1.SG	
1.	*mæn*	***ʔin***	***pesær-o***	*mi-bin-æm.*	PERSIAN
	I	this	boy-SP	IND-see-1.SG	
2.	*mæn*	***ʔin***	***doxtær-o***	*mi-bin-æm.*	
	I	this	girl-SP	IND-see-1.SG	

It can be inferred from the data above that, like agreement in number between

demonstratives and head nouns, the agreement in gender between demonstratives and head nouns yields the same results. The parameter values for Kırdki in both nominative and accusative cases are positive, and the parameter value in the accusative case for Kurmanji is positive. The values of the parameters in all other varieties are negative.

3.1.5 Grammaticalized Agreement in Gender Between Adjectives and Head Nouns in the Nominative Case

This kind of agreement differentiates languages with regard to gender agreement between adjectives and head nouns in the nominative case. If the distinction is established in a language variety, the parameter value for that variety will be positive. If it does not, the value will be negative. Consider the following examples:

Ex: 1. a **tall boy** sees me. ENGLISH
2. a **tall girl** sees me.

1.	***kʊr-æk-i***	***dɪrezʒ***	*mɪn*	*dɪ-bin-æ.*	KURMANJI
	boy-INDF-LK.M	tall	me	IND-see-3.SG	
2.	***kætʃ-æk-a***	***dɪrezʒ***	*mɪn*	*dɪ-bin-æ.*	
	girl-INDF-LK.F	tall	me	IND-see-3.SG	
1.	***kʊr-ek-i***	***dɪrezʒ***	*mɪn*	*ʔæ-bin-e.*	SORANI
	boy-INDF-LK	tall	me	IND-see-3.SG	
2.	***kæniʃk-ek-i***	***dɪrezʒ***	*mɪn*	*ʔæ-bin-e.*	
	girl-INDF-LK	tall	me	IND-see-3.SG	
1.	***kʊr-e***	***dɪrezʒ-i***	*mɪ*	*dɪ-win-i.*	KALHORI
	boy-LK	tall-INDF	me	IND-see-3.SG	
2.	***dɪwæt-e***	***dɪrezʒ-i***	*mɪ*	*dɪ-win-i.*	
	girl-LK	tall-INDF	me	IND-see-3.SG	
1.	***kʊr-ewæ***	***dɪrezʒ***	*ʔæmn-i*	*mæ-win-o.*	GORANI
	boy-INDF	tall	me-OBL	IND-see-3.SG	
2.	***kɪnatʃ-ewæ***	***dɪrezʒ-æ***	*ʔæmn-i*	*mæ-win-o.*	
	girl-INDF	tall-INDF.F	me-OBL	IND-see-3.SG	
1.	***laʒæk-end-o***	***dærg***	*mɪ*	*vin-æn-o.*	KIRDKI
	boy-INDF-M	tall	me	see-IND-3.SG	
2.	***kæjæk-end-a***	***dærg-æ***	*mɪ*	*vin-æn-a*	
	girl-INDF-F	tall-F	me	see-IND-3.SG	
1.	***kʊr-e***	***dɪriʒ-i***	*mɪn-æ*	*mɪ-wn-i.*	LAKI
	boy-LK	tall-INDF	me-OBL	IND-see-3.SG	
2.	***dɪt-e***	***dɪriʒ-i***	*mɪn-æ*	*mɪ-wn-i.*	
	boy-LK	tall-INDF	me-OBL	IND-see-3.SG	
1.	***kʊr-e***	***deraz-i***	*mo-ne*	*ʔi-bin-e.*	LORI
	boy-LK	tall-INDF	me-OBL	IND-see-3.SG	
2.	***dowær-e***	***deraz-i***	*mo-ne*	*ʔi-bin-e.*	
	boy-LK	tall-INDF	me-OBL	IND-see-3.SG	
1.	***pesær-e***	***deraz-i***	*mæn-o*	*mi-bin-e.*	PERSIAN
	boy-LK	tall-INDF	me-SP	IND-see-3.SG	
2.	***doxtær-e***	***deraz-i***	*mæn-o*	*mi-bin-e.*	
	boy-LK	tall-INDF	me-SP	IND-see-3.SG	

3.1.6 Grammaticalized Agreement in Gender Between Adjectives and Head Nouns in the Accusative Case

This type of agreement differentiates languages with regard to gender agreement between adjectives and head nouns in the accusative case. If the distinction is established in a language variety, the parameter value for that variety will be positive. If it does not, the value will be negative. Consider the following examples:

Ex: 1. I see a **tall boy**. ENGLISH
2. I see a **tall girl**.

1.	*ʔæz*	***kʊr-æk-i***	***dɪreʒ***	*dɪ-bin-ɪm.*	KURMANJI
	I	boy-INDF-LK.M	tall	IND-see-1.SG	
2.	*ʔæz*	***kætʃ-æk-a***	***dɪreʒ***	*dɪ-bin-ɪm.*	
	I	girl-INDF-LK.F	tall	IND-see-1.SG	
1.	*mɪn*	***kʊr-ek-i***	***dɪreʒ***	*ʔæ-bin-ɪm.*	SORANI
	I	boy-INDF-LK	tall	IND-see-1.SG	
2.	*mɪn*	***kɪtʃ-ek-i***	***dɪreʒ***	*ʔæ-bin-ɪm.*	
	I	girl-INDF-LK	tall	IND-see-1.SG	
1.	*mɪ*	***jej kʊr-e***	***dɪriʒ***	*dɪ-win-ɪm.*	KALHORI
	I	a boy-SP	tall	IND-see-1.SG	
2.	*mɪ*	***jej dɪwæt-e***	***dɪriʒ***	*dɪ-win-ɪm.*	
	I	a boy-SP	tall	IND-see-1.SG	
1.	*ʔæm*	***kʊr-ewæ***	***dɪreʒ-i***	*mæ-win-u.*	GORANI
	I	boy-INDF	tall-OBL.M	IND-see-1.SG	
2.	*ʔæm*	***kɪnatʃ-ewæ***	***dɪreʒ-æ***	*mæ-win-u.*	
	I	girl-INDF	tall-OBL-F	IND-see-1.SG	
1.	*ʔæz*	***laʒæk-end-o***	***derg-i***	*vin-æn-o.*	KIRDKI
	I	boy-INDF-LK.M	tall-OBL.M	see-IND-1.SG	
2.	*ʔæz*	***kæjnæk-ed-a***	***dærg-æ***	*vin-æn-a.*	
	I	boy-INDF-LK.F	tall-OBL.F	see-IND-1.SG	
1.	*mɪ*	***kʊr-e***	***dɪriʒ-i***	*mɪ-wn-ɪm.*	LAKI
	I	boy-LK	tall-OBL	IND-see-1.SG	
2.	*mɪ*	***dɪt-e***	***dɪriʒ-i***	*mɪ-wn-ɪm.*	
	I	girl-LK	tall-OBL	IND-see-1.SG	
1.	*mo*	***kʊr-e***	***deraz-i-ne***	*ʔi-bin-om.*	LORI
	I	boy-LK	tall-INDF-OBL	IND-see-1.SG	
2.	*mo*	***dowær-e***	***deraz-i-ne***	*ʔi-bin-om.*	
	I	girl-LK	tall-INDF-OBL	IND-see-1.SG	
1.	*mæn*	***pesær-e***	***deraz-i***	*mi-bin-æm.*	PERSIAN
	I	boy-LK	tall-OBL	IND-see-1.SG	
2.	*mæn*	***doxtær-e***	***deraz-i***	*mi-bin-æm.*	
	I	girl-LK	tall-OBL	IND-see-1.SG	

The data above show that gender agreement between adjectives and head nouns is established in Kırdki and Gorani. There are no such agreements in the other varieties. The parameter values of gender agreement between adjectives in Kırdki and Gorani are positive, whereas they are negative in the other varieties.

3.2 Semantic Person

The parameter of semantic person differentiates languages that express distinctions of person in pronouns, whether personal or reflexive pronouns. This parameter distinguishes languages like Mandarin and Cantonese from Japanese. The following pronouns are investigated to determine whether this feature exists in Kurdish varieties and the Persian language.

Ex:

1.	I	myself	ENGLISH
2.	You	yourself	
3.	He	himself	
4.	We	ourselves	
5.	You	yourselves	
6.	They	themselves	
1.	ʔæz	bɪxwæ	KURMANJI
2.	tu	bɪxwæ	
3.	ʔæw	bɪxwæ	
4.	ʔæm	bɪxwæ	
5.	hun	bɪxwæ	
6.	ʔæw	bɪxwæ	
1.	mɪn	xo-**m**	SORANI
2.	to	xo-**t**	
3.	ʔæw	xo-**j**	
4.	ʔemæ	xo-**man**	
5.	ʔewæ	xo-**tan**	
6.	ʔæwan	xo-**jan**	
1.	mɪ	xwæ-**m**	KALHORI
2.	to	xwæ-**t**	
3.	ʔæw	xwe-**j**	
4.	ʔemæ	xwæ-**man**	
5.	ʔewæ	xwæ-**tan**	
6.	ʔæwan	xwæ-**jan**	
1.	ʔæm	wi-**m**	GORANI
2.	to	wi-**w**	
3.	ʔæd	wi-**ʃ**	
4.	ʔemæ	wi-**man**	
5.	ʃɪmæ	wi-**tan**	
6.	ʔæ̲di	wi-**ʃan**	
1.	ʔæz	bɪxo	KIRDKI
2.	tɪ	bɪxo	
3.	ʔa/ʔo	bɪxo	
4.	ma	bɪxo	
5.	ʃima	bɪxo	
6.	ʔe	bɪxo	

1.	mɪ	wɪʒ-**ɪm**	LAKI
2.	tü	wɪʒ-**ɪt**	
3.	ʔəw	wɪʒ-**i**	
4.	ʔemæ	wɪʒ-**mon**	
5.	homæ	wɪʒ-**ton**	
6.	ʔəwan	wɪʒ-**jon**	
1.	mo	xo-**m**	LORI
2.	to	xo-**t**	
3.	ho	xo-**s**	
4.	ʔima	xo-**mon**	
5.	ʔisa	xo-**ton**	
6.	ʔono	xo-**son**	
1.	mæn	xod-**æm**	PERSIAN
2.	to	xod-**et**	
3.	ʔun	xod-**eʃ**	
4.	ma	xod-**emun**	
5.	ʃoma	xod-**etun**	
6.	ʔuna	xod-**eʃun**	

The data above reveal that the distinction of person can be observed in Sorani, Kalhori, Gorani, Laki, Lori, and Persian. However, this kind of distinction is not observed in Kurmanji and Kırdki. So, the parameter values for this parameter are negative in Kurmanji and Kırdki.

3.3 Semantic Gender

The parameter of semantic gender differentiates languages that contrast two third person singular pronouns encoding perceived biological sex and/or animacy (e.g., English) from languages that do not encode animacy or biological sex (e.g., Finnish, Turkish, and Hungarian). To obtain a finer-grained analysis of the considered language varieties, the distinctions of gender and animacy in third person singular pronouns are analyzed in sections 3.3.1 and 3.3.2, respectively.

3.3.1 The Overt Distinction of the Gender of Third Person's Pronoun

The distinction is established in a language variety if there is an overt distinction of gender in the third person singular pronouns. If a language variety encodes the gender in the third person singular pronouns, the parameter value for that variety will be positive. If it does not, the value will be negative. Consider the following examples:

Ex: 1. I know **Azad**; **he** is very clever. ENGLISH
2. I know **Rojan**; **she** is very clever.

1.	*ʔæz*	***Azad***	*nas*	*dɪ-k-ɪm;*	***ʔæw***	*pɪrr*	*ziræk-æ.*	KURMANJI
	I	Azad	know	IND-do-1.SG	he	very	clever-is	
2.	*ʃæz*	***Rojan***	*nas*	*dɪ-k-ɪm;*	***ʔæw***	*pɪrr*	*ziræk-æ.*	
	I	Rojan	know	IND-do-1.SG	she	very	clever-is	
1.	*mɪn*	***Azad***	*dæ-nas-ɪm;*		***ʔæw***	*zor*	*ziræk-æ.*	SORANI
	I	Azad	IND-know-1.SG		he	very	clever-is	

2.	*mɪn*	***Rojan***	*dœ-nas-ɪm;*	***ʔæw***	*zor*	*zirœk-œ.*	
	I	Rojan	IND.know-1.SG,	she	very	clever-is	
1.	*mɪn*	***Azad***	*ʃinas-ɪm;*	***ʔæw***	*fɪrœ*	*zɪrɪng-œ*	KALHORI
	I	Azad	know-1.SG	he	very	clever-is	
2.	*mɪn*	***Rozan***	*ʃinas nas-ɪm;*	***ʔæw***	*fɪrœ*	*zɪrɪng-œ*	
	I	Rojan	know-1.SG	she	very	clever-is	
1.	*ʔæmɪn*	***Azad-i***	*mœ-ʒnas-u;*	***ʔado***	*fɪrœ*	*zirœk-œn*	GORANI
	I	Azad-OBL	IND-know-1.SG	he	very	clever-3.SG	
2.	*ʔæmɪn*	***Rozan-i***	*mœ-ʒnas-u;*	***ʔadœ***	*fɪrœ*	*zirœk-œn-e*	
	I	Rojan-OBL	IND-know-1.SG	she	very	clever-3.SG	
1.	*ʔæz*	*Azad-i*	*ʃinas-nœn-a.*	***ʔo***	*zafzirœk-o.*		KIRDKI
	I	Azad-OBL	know-IND-3rd.SG		he very	clever-is	
2.	*ʔæz*	*Azad-i*	*ʃinas-nœn-a.*	***ʔa***	*zafzirœk-o.*		
	I	Azad-OBL	know-IND-3rd.SG	he	very	clever-is	
1.	*mɪ*	***Azad-œ***	*mœ-ʃnas-ɪm;*	***ʔəw***	*fɪrœ*	*zɪrɪŋ-œ*	LAKI
	I	Azad-OBL	IND-know-1.SG	he	very	clever-COP.3.SG	
2.	*mɪ*	***Rozan-œ***	*mœ-ʃnas-ɪm;*	***ʔəw***	*fɪrœ*	*zɪrɪŋ-œ*	
	I	Rojan-OBL	IND-know-1.SG	she	very	clever-COP.3.SG	
1.	*mo*	***Azad-ene***	*ʔi-ʃnas-om;*	***ho***	*xejli*	*zebel-e.*	LORI
	I	Azad-OBL	IND-know-1.SG	he	very	clever-COP.3.SG	
2.	*mo*	***Rozan-ne***	*ʔi-ʃnas-om;*	***ho***	*xejli*	*zebel-e.*	
	I	Rojan-OBL	IND-know-1.SG	she	very	clever-COP.3.SG	
1.	*mœn*	***Azad-o***	*mi-ʃnas-œm;*	***ʔun***	*xejli*	*zerœng-e.*	PERSIAN
	I	Azad-OBL	IND-know-1.SG	he	very	clever-COP.3.SG	
2.	*mœn*	***Rojan-o***	*mi-ʃnas-œm;*	***ʔun***	*xejli*	*zerœng-e.*	
	I	Rojan-OBL	IND-know-1.SG	she	very	clever-COP.3.SG	

The overt gender distinction of the third person pronoun is observed only in Gorani and Kırdki. Therefore, the parameter value in Gorani and Kırdki is positive and it is negative in the other language varieties.

3.3.2 The Overt Distinction Animacy of Third Person's Pronoun

The distinction is established in a language variety if there is an overt distinction of animacy in the third person singular pronouns. If a language variety encodes the animacy in the third person singular pronouns, the parameter value in that variety will be positive. If it does not, the value will be negative. Consider the following examples:

Ex: I have a **house**; **it** is very big. ENGLISH

1.	***mal-a***	*mɪn*	*hœjœ;*	***Ø***	*pɪrr*	*mœzɪn-œ.*	KURMANJI
	house-LK	my	COP.SG	(it)	very	big -is	
1.	*mɪn*	***mał****-ek-ɪm*	*hœjœ;*	***Ø***	*zor*	*gœwrœ-jœ.*	SORANI
	******mɪn*	***mał****-ek-ɪm*	*hœjœ;*	***æw***	*zor*	*gœwrœ-jœ.*	
	I	house-INDF-1.SG	have	(it)	very	big-is	

1.	*mɪn*	***mal-ig***	*dɪr-em;*	***Ø***	*firæ*	*gæwra-s.*	KALHORI
	**mɪn*	***mal-ig***	*dɪr-em;*	*æw*	*firæ*	*gæwra-s.*	
	I	house-DEF	have-1.SGit		very	big-is	
1.	*ʔæmɪn*	***jane-wæ-m***	*hæn;*	***Ø***	*firæ*	*goræ-n.*	GORANI
	I	house-INDF-CL.1.SG	have	(it)	very	big-3.SG	
1.	*kæjæ-je*	*mɪ*	*æst-o.*	***Ø***	*zafgɪrd-o.*		KIRDKI
	house-LK	my	is-3.SG	(it)	very	bid-is	
1.	*mɪ*	***mal-i***	*dɪr-ɪm;*	***Ø***	*firæ*	*kælɪn-æ.*	LAKI
	I	house-DEF	have-1.SG	(it)	very	big-COP.3.SG	
1.	*mo*	*je hone*	*dar-om;*	***Ø***	*xejli*	*gæp-e.*	LORI
	I	one house	have-1.SG	(it)	very	big-COP.3.SG	
1.	*mæn*	*je xune*	*dar-æm;*	***Ø***	*xejli*	*bozorg-e.*	PERSIAN
	I	one house	have-1.SG	(it)	very	big-COP.3.SG	

The overt animacy distinction of the third person pronoun is not realized in any language variety. Therefore, the parameter value in all varieties is negative. To provide a more complete picture of all the parametric values for the language varieties, the values of 49 parameter manifestations are presented in Figure 1.

	Parameters	**Kurmanji**	**Sorani**	**Kalhori**	**Gorani**	**Kırdki**	**Laki**	**Lori**	**Persian**
1	grammaticalized morphology	+	+	+	+	+	+	+	+
2	±Person Agr on Dem-Poss	-	-	-	-	-	-	-	-
3	±Person Agr on Q-Poss	-	-	-	-	-	-	-	-
4	±Person Agr on Art-Poss	-	-	-	-	-	-	-	-
5	±Person Agr on Poss-N	-	-	-	-	-	-	-	-
6	±Person Agr on Adj-Poss	-	-	-	-	-	-	-	-
7	±Num Agr on Dem-N (Nom)	−	−	−	−	+	−	-	-
8	±Num Agr on Dem-N (Acc)	+	-	-	−	+	−	-	-
9	±Num Agr on Q-N	+	+	+	+	+	+	+	+
10	±Num Agr on Art-N	-	−	−	−	−	-	-	-
11	±Num Agr on Poss-N	-	-	-	-	-	−	-	-
12	±Num Agr on Adj-N (Nom)	-	−	−	+	+	-	-	-
13	±Gender Agr on Dem-N (Nom)	−	−	−	−	+	−	-	-
14	±Gender Agr on Dem-N (Acc)	+	−	−	−	+	−	-	-
15	±Gender Agr on Q-N	−	−	−	−	−	−	-	-
16	±Gender Agr on Art-N	−	−	−	+	−	−	-	-
17	±Gender Agr on Poss-N	−	−	−	−	−	−	-	-
18	±Gender Agr on Adj-N (NoM)	−	−	−	+	+	−	-	-
19	±Gender Agr on Adj-N (Acc)	−	−	−	+	+	−	-	-
20	±Variable Case on Pronouns	+	+	−	+	+	+	+	-
21	±Variable Case on Demonstratives	+	-	-	-	+	-	-	-

22	±Variable Case on Quantifiers	-	-	-	-	-	-	-	-
23	±Variable Case on Articles	-	-	-	-	-	-	-	-
24	±Variable affixed Case on Nouns	+	+	-	+	+	+	+	-
25	±Variable Form of Case on Nouns	+	-	-	-	+	-	-	-
26	±morphological Agr on Verb/Argument	+	+	+	+	+	+	+	+
27	±Overt expletive in Subject Position	-	-	-	-	-	-	-	-
28	±Person-marking Pronouns before Adjs	+	+	+	+	+	+	+	+
29	±Marked Maximal N Argument	-	+	+	+	-	+	+	-
30	± semantic Person	-	+	+	+	-	+	+	+
31	*Number Bound morpheme on arguments (Nom)*	-	+	+	+	+	+	+	+
32	*Number Bound morpheme on arguments (Acc)*	+	+	+	+	+	+	+	+
33	*Number Agr on Argument and Pres Intr Verb*	+	+	+	+	+	+	+	+
34	*Number Agr on Argument and PST Intr Verb*	+	+	+	+	+	+	+	+
35	*Number Agr on Argument and PST Intr Verb*	+	+	+	+	+	+	+	+
36	*Number Agr on Argument and PST Trs Verb*	-	-	+	-	-	-	+	+
37	Number Agr on N and Adjs (Nom)	-	-	-	-	+	-	-	-
38	Number Agr on N and Adjs (Acc)	-	-	-	-	+	-	-	-
39	Number Agr on 3rd Pronoun and its antecedent	-	+	+	+	-	+	+	+
40	±N with variable Number Morphology (Nom)	-	+	+	+	+	+	+	+
41	±N with variable Number Morphology (Acc)	+	+	+	+	+	+	+	+
42	± Bare Noun in Argument Function	-	+	+	+	+	+	+	+
43	Systematic realization of Number on N (Nom)	-	+	+	+	+	+	+	+
44	Systematic realization of Number on N (Acc)	+	+	+	+	+	+	+	+
45	±Gender Agr on Dem-N (Nom)	-	-	-	-	-	-	-	-
46	±Gender Agr on Dem-N (Acc)	+	-	-	-	+	-	-	-
47	±Gender Agr on Adj-N	-	-	-	+	+	-	-	-
48	± overt distinction gender of 3rd person Pronoun	-	-	-	+	+	-	-	-
49	± overt distinction Animacy of 3rd person Pronoun	-	-	-	-	-	-	-	-

Figure 1. The List of Some Parameters and Parameter Values in the Domain of Determinet Phrase (DP) in Kurdish Varieties and Persian

To calculate the parametric distances, the number of differences between each pair is divided by the sum of the differences and identities. For instance, according to Figure 1, the number of differences (d) between Kurmanji and Sorani is 12, and the number of identities (i) is also 12. To calculate the distance, the number of differences, 12, is divided by the sum of

the differences and identities, 12 + 12 = 24. The result is 0.50. This number indicates the syntactic (parametric) distance between the pair varieties of Kurmanji and Sorani as follows:

Kurmanji vs Sorani
The number of differences (d) = 12
The number of identities (i) = 12
The sum of the differences and identities (d + i) = 24

$$\text{Distance}_{\text{Kur-Sor}} = \frac{d}{d+i} = \frac{12}{12+12} = \frac{12}{24} = 0.50$$

The syntactic (parametric) distances between all pairs of language varieties are calculated using the procedure described above and are indicated in Figure 2.

	Kurmanji							
Kurmanji	0	Sorani						
Sorani	0.50	0	Kalhori					
Kalhori	0.60	0.15	0	Gorani				
Gorani	0.60	0.24	0.35	0	Kirdki			
Kırdki	0.60	0.52	0.59	0.42	0	Laki		
Laki	0.50	0.00	0.15	0.09	0.54	0	Lori	
Lori	0.56	0.05	0.10	0.27	0.52	0.05	0	Persian
Persian	0.56	0.20	0.06	0.39	0.57	0.20	0.15	0

Figure 2. Matrix of Parametric (Syntactic) Distance

By feeding the pairwise distances obtained from the set of characters in Figure 2 into the UPGMA distance-based phylogenetic algorithm, the following tree is generated (Figure 3).

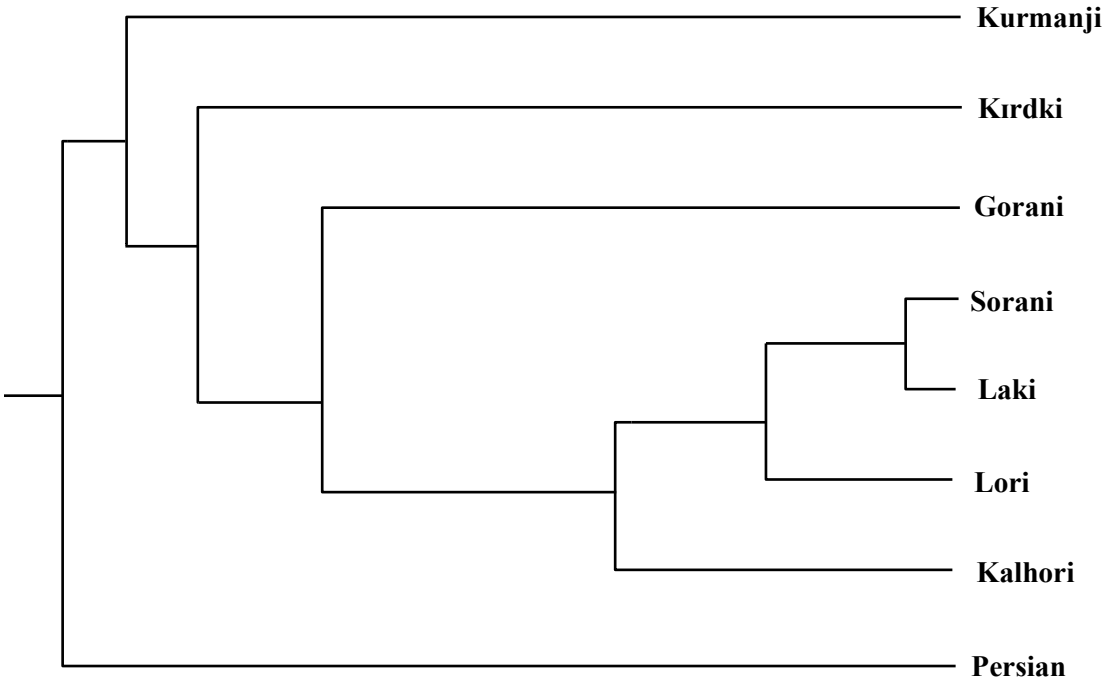

Figure 3. UPGMA Tree Achieved From the Syntactic Distances in Figure 2 of Kurdish and Other Related Varieties

The tree keeps Kurmanji and Kırdki far from the other varieties and, interestingly, groups Sorani and Laki together as two varieties with no differences. Lori is at a close distance compared to Sorani, Laki, and Kalhori. Even though it was assumed that Persian would be far from these varieties, it shows some relationships with them.

The most distinctive features in the language varieties are gender and some forms of person and number agreement. In Kurmanji, Kırdki, and Gorani, gender features are observed in the DP structure, but gender is not observed in Sorani, Kalhori, Laki, Lori, and Persian. Regarding person agreement, there is no agreement in person between reflexive pronouns and their antecedents in Kurmanji and Kırdki, while it is observed in the other language varieties. With respect to number agreement, Kurmanji interestingly indicates no number marking on nouns in the nominative case; however, it appears in the accusative case.

In this way, the syntactic distances and relationships between language varieties are presented, in which some irregularities can be seen. One of the main points regarding the relationships concerns the classical classifications of the Kurdish dialects, including northern Kurdish (Kurmanji), central Kurdish (Sorani), and southern Kurdish (Kalhori). This classification is not confirmed by the results of this study. Laki and Sorani are in a close syntactic relationship, while Kurmanji and Kalhori are in a far distance from Sorani. An interesting point is that Lori is in a closer relationship with Sorani in comparison with Kurmanji and Kalhori.

4. Conclusion

This study, using the Parametric Comparison Method as a tool to measure the syntactic distances between Kurdish language varieties and Persian, revealed some interesting points.

Firstly, the parametric comparison can successfully measure the distances between genealogically related languages and indicate the fine-grained similarities and differences more obvious than the classical comparative methods.

Secondly, the results of this study do not support the claim that there are only three dialects of the Kurdish language, including northern Kurdish (Kurmanji), central Kurdish (Sorani), and southern Kurdish (Kalhori). There are language varieties that show close relationships to these three varieties, such as Laki, Lori, and Kırdki. Since the approach used in this study claims linguistic parameters are robust indicators of genealogical relations, the close relationship of Lori and Laki should be considered important.

Thirdly, Kalhori is at a close distance from the Persian language in comparison with the other language varieties. Therefore, the Persian language is at one end of the spectrum and Kurmanji is at the other end.

Ethical Considerations

Not applicable

Funding

Not applicable

Conflict of Interest

The authors declare no conflicts of interest related to this research.

References

Baker, M. C. (2008). The atoms of language: The mind's hidden rules of grammar. Basic books.

Bromham, L., & Penny, D. (2003). The modern molecular clock. *Nature Reviews Genetics*, *4*, 216–224. https://doi.org/10.1038/nrg1020

Ceolin, A., Guardiano, C., Irimia, M. A., & Longobardi, G. (2020). Formal syntax and deep history. *Frontiers in Psychology*, *11*, Article 488871. https://doi.org/10.3389/fpsyg.2020.488871

Chomsky, N. (1955). Logical syntax and semantics: Their linguistic relevance. *Language*, *31*(1), 36–45.

Chomsky, N. (1957a). Logical structures in language. *Journal of the American Society for Information Science*, *8*(4), 284–291. https://doi.org/10.1002/asi.5090080406

Chomsky, N. (1981b). Principles and parameters in syntactic theory. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics* (pp. 32–75). Longman.

Crisma, P., Guardiano, C., & Longobardi, G. (2020). Syntactic diversity and language learnability. *Studi e Saggi Linguistici*, *58*(2), 99–130.

Croft, W. (2008). Evolutionary linguistics. *Annual Review of Anthropology*, *37*(1), 219–234. https://doi.org/10.1146/annurev.anthro.37.081407.085156

De Gregorio, J., Toral, R., & Sánchez, D. (2024). Exploring language relations through syntactic distances and geographic proximity. *arXiv*. https://doi.org/10.48550/arXiv.2403.18430

Dunn, M. (2015). Language phylogenies. In B. B. Bowern & B. Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 190–211). Routledge.

Felsenstein, J. (2004). *Inferring phylogenies.* Sinauer Associates.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (Vol. 2, pp. 73–113). MIT Press.

Greenberg, J. H. (1987). *Language in the Americas*. Stanford University Press.

Greenberg, J. H. (2000). Indo-European and its closest relatives: The Eurasiatic language family: Vol. 1. Grammar. Stanford University Press.

Guardiano, C., & Longobardi, G. (2005). Parametric comparison and language taxonomy. In M. Batllori, M. L. Hernanz, C. Picallo, & F. Roca (Eds.), *Grammaticalization and parametric variation* (pp. 149–174). Oxford University Press.

Hale, M. (2007). Historical linguistics: Theory and method. Blackwell.

Lightfoot, D. (1999). The development of language: Acquisition, change, and evolution. Wiley-Blackwell.

Longobardi, G. (2001). The structure of DPs: Some principles, parameters, and problems. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 562–603). Blackwell.

Longobardi, G. (2003). Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook*, *3*(1), 101–138.

Longobardi, G., Ghirotto, S., Guardiano, C., Tassi, F., Benazzo, A., Ceolin, A., & Barbujani, G. (2015a). Across language families: Genome diversity mirrors linguistic variation within Europe. *American Journal of Physical Anthropology*, *157*(4), 630–640. https://doi.org/10.1002/ajpa.22758

Longobardi, G., & Guardiano, C. (2009). Evidence for syntax as a signal of historical relatedness. *Lingua*, *119*(11), 1679–1706. https://doi.org/10.1016/j.lingua.2008.09.012

Longobardi, G., Guardiano, C., Silvestri, G., Boattini, A., & Ceolin, A. (2013). Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics*, *3*(1), 122–152. https://doi.org/l0.1075/jhl.J.L07lon

Longobardi, G., & Treves, A. (2023). Grammatical Parameters from a Gene-like Code to Self-Organizing Attractors. *arXiv*. https://doi.org/10.48550/arXiv.2307.03152

Matras, Y., Haig, G., & Öpengin, E. (2022). *Structural and typological variation in the dialects of Kurdish*. Springer Nature. https://doi.org/10.1007/978-3-030-78837-7

Orr, R. A. (1999). Evolutionary biology and historical linguistics. *Diachronica*, *16*(1), 123–157.

Roberts, I. G. (1993). *Verbs and diachronic syntax: A comparative history of English and French* (Studies in Natural Language and Linguistics Theory, Vol. 28). Springer Science & Business Media.

Roberts, I. G. (1998). [Review of the book Historical syntax in cross-linguistic perspective, by A. C. Harris & L. Campbell]. *Romance Philology*, *51*(3), 363–370. http://www.jstor.org/stable/44941034

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationship. *University of Kansas Science Bulletin*, *38*(22), 1409–1438.

Windfuhr, G. (2009). *The Iranian languages*. Routledge.