

**Citation:** Pericliev, V. (2017). Kin term patterns and language families. *Cognitive Studies / Études cognitives*, 2017(17). <https://doi.org/10.11649/cs.1374>

VLADIMIR PERICLIEV

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria  
[peri@math.bas.bg](mailto:peri@math.bas.bg)

## KIN TERM PATTERNS AND LANGUAGE FAMILIES

### Abstract

The anthropologist G. P. Murdock has found a strong correlation between the kin term patterns (or feature-values) for the relative (feature) “sibling” and language families. This important finding for language classification, however, has not been pursued further. In particular, it has not yet been tested whether the kin term patterns domain as a whole, including the patterns for other features (“grandparents”, “uncles”, “aunts”, “nephews and nieces”, etc.), is sufficient to demarcate all language families from one another. This paper presents a large-scale computational profiling of all language families in terms of their kin term patterns. The most significant findings are: (i) that language families can be quite neatly differentiated on the basis of their kin term patterns, and therefore these patterns may be considered as strong indicators of genetic affiliation, and (ii) that the kin term patterns for the features “nephews and nieces (= siblings’ children)”, “siblings”, and “siblings-in-law” — i.e. all features including the idea of siblings — are the best predictors of genetic affiliation, as they are significantly more frequently used in the profiles than any other feature.

**Keywords:** kin term patterns; language classification; computational classification

## 1 Introduction

Kinship terminological systems have attracted a lot of attention both from linguists and from anthropologists. Despite various (often opposing) interpretations and different methodological approaches, a large body of terminological systems of the world’s languages has been collected over the years. The kin terms domain thus provides a solid empirical basis for comparative linguistic and anthropological work.

A major contribution of anthropologists is the concept of so called “kin term patterns” for the different types of relatives, or kinsmen: grandparents, grandchildren, uncles, aunts, nephews and nieces, siblings, cross-cousins, and siblings-in-law. These types of relatives are usually referred to as *features*, and the kin term patterns for these relatives are referred to as *values* of these features.

The famous anthropologist G. P. Murdock (Murdock, 1968) has found a strong correlation between kin term patterns for the feature “siblings” and language families. This is an important discovery for historical linguistics, as it shows the historical stability of the feature, and suggests

the hypothesis that kin term patterns as a whole may serve to demarcate all language families from one another, a hypothesis that has not been tested so far. This paper presents a large-scale computational profiling of the world language families and briefly discuss the results of this test.

Section 2 introduces the concept of kin term patterns and Murdock’s findings. Section 3 formulates the problem that is to be examined. Section 4 describes Murdock’s database of the kin term patterns of the world’s languages, which constitutes the empirical base for the paper’s computations. Section 5 outlines the computational mechanism used to accomplish the profiling of language families in terms of kin term patterns, and Section 6 discusses the results of the profiling.

## 2 Kin term patterns and Murdock’s findings

*Kin term patterns* reveal the semantic structuring of kin terms, or how the space of relatives is partitioned by the kin terms in a language. For example, for the relative grandparent, some languages may use a “Bisexual Pattern” (having two terms, distinguished by sex, which can be glossed as “grandfather” and “grandmother”), others a “Merging Pattern” (having a single undifferentiated term, which can be glossed as “grandparent”), and so on. The idea of kin term patterns has its roots in the pioneering work of Louis Henry Morgan (1870) and was later developed by other eminent scholars such as Boas and Kroeber. The basic kin term patterns were developed by Lowie (1928) and Kirchhoff (1932) for avuncular and nepotic terms and by Spier (1925) and Murdock (1949) for cousin terminology.

An important work relevant to the problem discussed in this article is that of Murdock (1968) on sibling terminological patterns. In trying to classify the most frequently used patterns for siblings in 800 societies worldwide, he unexpectedly discovered a strong correlation of sibling patterns with language families and sub-families, a result that was to be re-confirmed later by Marshall (1984) on data from a single linguistic stock, encompassing 223 languages from islands in Oceania.

Murdock isolates seven basic sibling patterns and ascertains that the distribution of these patterns closely follows the boundaries of the then accepted language families and sub-families, in the sense that each such linguistic division uses one numerically prevalent pattern. Only in two world regions is the distribution of sibling patterns based on areal principles, overriding the linguistic divisions. The first region includes most of north and northeast Africa, all of Europe, and a large part of southwest Asia. The second region includes the area of higher civilization in Asia, including societies speaking Dravidian, Japanese, the Indic sub-family of Indo-European, some speakers of Tibeto-Burman, and the Altaic sub-families Turkic and Tungusic. Since there are seven basic sibling kin term patterns, the employment of a single pattern, especially in large families comprising many languages, is an indicator of the non-random nature of this distribution.

By way of illustration of Murdock’s findings, the “European Pattern” (or “Brother-Sister Pattern”), characterized by single distinction in sex, as in the English *brother* and *sister*, is found in large or medium-sized language families (Afro-Asiatic, Indo-European, Sudanic, Uralic), in small families (e.g. Basque, Georgian, Yuki, Ket), and in sub-families (Southern Khoisan). In a similar fashion, Murdock reviews the distribution, according to linguistic divisions, of the remaining sibling patterns. He also discusses the testing, in the light of his results, of proposed groupings for Austronesian by Isidore Dyen, as well as some interesting cases where cultural and linguistic diffusion might be the driving force for the sibling pattern distribution, rather than genetic affiliation.

## 3 Formulation of the problem

Murdock’s paper is an important contribution to the understanding of the relationship between kin term patterns and language classification, in that it provides the insight that sibling terminology is highly relevant for linguistic grouping. This finding implies both the historical stability of the feature “Sibling” and its individualizing nature regarding languages and language families. The

study of the individualizing nature of certain features in languages and language families is a trend that has a significant history in linguistic research. In an historical perspective, the individualizing approach has typically emphasized the global characteristics of a language as forming an organic unity. Sapir, for instance, talks about a “basic plan, a certain cut”, or a “structural genius” of each language, or the presence of a combination of features, which is much more fundamental than the presence of each individual feature in isolation. This overall genius of a language, in the Humboldtian and Whorfian linguistic traditions, was believed to have echoes in the culture and national character of the speakers of the language.

The general problem that this paper is focusing on is whether the domain of kin term patterns can serve as an individualizing feature of language families. This problem is three-fold and includes three important interrelated questions:

- I *The potential of kin term patterns for language family discrimination* (or, can we discriminate all the known language families in terms of the kin term patterns domain?)
- II *The contribution of features to language family discrimination* (or, what are the overall features that are necessary and sufficient for this discrimination and what is the contribution of each feature in singling out each language family? Is the feature “Sibling”, proposed by Murdock, the best predictor of affiliation, or are other features more predictive?)
- III *The idiosyncrasy of language families* (or, how do language families differ with respect to the number of features they need to be singled out, or are some families more easily discriminated than others?)

## 4 Murdock’s kin term patterns database

This investigation is based on Murdock (1970), a database containing the patterns for all relatives, or features.

Murdock describes the terminological classification systems of 566 languages (“societies” in his terminology) from 194 of the 200 cultural provinces that he had isolated. This dataset is the most representative compilation of kinship terminologies to date. He justly says that “This collection makes it possible — perhaps for the first time in the history of anthropology — to analyse important segments of culture on the basis of an examination of the total universe of described cultures, not merely of a sample or partial survey of such” (Murdock, 1970, p. 165). The data includes virtually all the systems published for Africa and aboriginal North and South America, and is only slightly less exhaustive for Eurasia and Oceania. Moreover, the dataset is based on files of over a thousand complete systems; the published dataset includes only those systems which differ from the remaining systems within the same sampled province in order to avoid duplication.

The Murdock dataset focuses on eight types of kin (or features):

1. Grandparents (GrPa),
2. Grandchildren (GrCh),
3. Uncles (PaBr),
4. Aunts (PaSi),
5. Nephews and nieces (male speaker) (SbCh),
6. Siblings (Sibl),
7. Cross-Cousins (CrCo),

## 8. Siblings-in-law (Sb-II).

Every type of kin (feature) is described in terms of kin term patterns (feature-values), showing the number of kin terms used for that kin, as well as their range of reference. The features are multi-valued and may have a dozen or more values.

It is important to note that Murdock’s data occasionally contains *uncertain* patterns (e.g. for some languages GrPa=A? would mean that the correct pattern is probably A, but this is not definitive) or *unknown*, missing, patterns (e.g. GrPa=? would mean that the pattern for grandparents for some language is unknown). In the processed dataset, the uncertainty in the patterns was ignored, basing the analysis on the pattern suggested as probable (and the question mark juxtaposed to a pattern was simply removed). The treatment of unknown patterns will be discussed in the next section.

For the purposes of this study, it was necessary to re-organise Murdock’s sample, which grouped the individual languages under “cultural provinces”, into a database which groups them under language families (phyla). To this end, the *Ethnologue* was used. This is a standard and constantly updated reference on world languages and language families. In the process of reorganising the data, omitted from consideration were (1) unclassified languages, (2) some isolate languages, (3) creoles, and (4) languages that would serve as a test bed for testing the predictions of the demarcation. This procedure led to discarding some 72 languages, so the total number of languages have actually processed is 494, which is still a very representative corpus.

Listed below are the language families, distributed in four large geographical areas. The numbers of attendant languages for each family are given in brackets.

*Area I. North and Central America:* Algic (16), Caddoan (5), Coahuiltecan (1), Eskimo-Aleut (4), Gulf (2), Hokan (14), Huavean (1), Iroquoian (2), Keres (1), Kiowa Tanoan (5), Mayan (5), Mixe-Zoque (1), Muskogean (2), Na-Dene (20), Oto-Manguean (4), Penutian (5), Salishan (9), Siouan (9), Totonacan (1), Uto-Aztecan (14), Wakashan (4), Yuki (1).

*Area II. South America and the Caribbean:* Araucanian (1), Arawakan (11), Aymaran (1), Barbacoan-Paezan (2), Carib (8), Chibchan (6), Choco (1), Chon (1), Jivaroan (1), Macro-Ge (8), Mataco-Guaicuru (4), Nambiquaran (1), Salivan (1), Tucanoan (2), Tupi (6), Uru-Chipaya (2), Witotoan (1), Yanomam (1).

*Area III. Africa, Europe and Asia:* Afro-Asiatic (20), Altaic (8), Andamanese (1), Austro-Asiatic (8), Austronesian (76), Basque (1), Chuk.-Kamchatkan (1), Daic (2), Dravidian (6), Hmong-Mien (1), Indo-European (28), Japanese (3), Khoisan (5), Niger-Congo (84), Nilo-Saharan (17), North Caucasian (1), Sino-Tibetan (18), South Caucasian (1), Uralic (2), Yukaghir (1).

*Area IV. Pacific:* Australian (6), East Papuan (1), Sepik-Ramu (3), Trans-New-Guinea (14).

## 5 Profiling language families in terms of kin term patterns

In order to answer the three questions raised in Section 3, it is necessary to discriminate all the language families from one another. This will show whether kin terms are predictively sufficient for the task and whether they will reveal the necessary and sufficient overall features, as well as reveal the set of discriminant features for each family. Given the latter, it is possible to compute the contribution of each feature to the demarcation of the families and compare the sizes of the discriminant sets.

To accomplish the task, the Maximally Parsimonious Discrimination program (MPD), was implemented. MPD is a computer program which can infer, given multiple classes with attendant instances of these classes (each instance being described by feature values), the *profiles* (= descriptions) of these classes, so that every class is differentiated from every other class by at least one feature value. MPD furthermore generates the most succinct class profiles, using the least number of overall features for the discrimination. In other words, each profile produced by MPD comprises the necessary and jointly sufficient feature values to demarcate each class from all remaining classes.

The program can also create so-called *derived* (or *complex*) *features* when no successful contrasts are achieved, which is not uncommon. The key idea of this feature creation is to express interactions between pairs from the given primitive features. Thus, two features M and N are combined to form the derived nominal feature M- $\times$ -N (see below). A detailed linguist-oriented description of the MPD program can be found in Pericliev (2010, Chapter 3), and a technical computational description in Valdes-Perez, Pereira and Pericliev (2000).

Each language family is associated with a set of instances (=individual languages), and each individual language is described in terms of features (=the set of 8 kin: GrPa, GrCh, PaBr, etc.) and values (=the respective kin term patterns for these features). The goal is to find a profile (=a set of feature values) for every language family, so that every family is demarcated from all remaining families by at least one feature-value.

A simple example may serve to illustrate contrasts in primitive and complex, or derived features. Listed below are the Australian and the Daic language families, described in terms of the eight features (kin) (this is the actual input to MPD). For ease of exposition, the values of features, or the kin term patterns, are designated by a capital letter, rather than their full name; e.g. the feature value pair GrPa = C stands for the *Bifurcate Bisexual Pattern* for grandparents, GrCh = C stands for the *Self-Reciprocal Pattern* for grandchildren and so on (an explanation and full listing of these patterns may be found in Murdock, 1970). A question mark “?” indicates that the pattern for the respective feature (=relative) is unknown.

Table 1: Contrasting Australian and Daic with primitive features

	<i>GrPa</i>	<i>GrCh</i>	<i>PaBr</i>	<i>PaSi</i>	<i>SbCh</i>	<i>Sibl</i>	<i>CrCo</i>	<i>Sb-il</i>
Australian								
Groote Eylandt	C	Y	A	B	B	A	D	E
Tiwi	C	N	A	B	B	A	B	E
Aranda	C	C	A	B	C	N	B	D
Dieri	C	C	A	B	A	D	B	E
Kariera	C	C	A	Ba	B	A	B	E
Murinbata	C	C	A	B	B	B	B	E
Daic								
Lao	N	A	L	K	?	A	C	?
Siamese	C	A	L	K	D	C	A	?

Two language families contrast with respect to a feature if they do not share a value (pattern) for that feature. In other words, the set of values for a feature in one family must not intersect with the set of values for the same feature in the other family. Thus, Australian and Daic contrast e.g. with respect to the feature GrCh (both Daic languages have the value A, while the Australian languages have Y, N or C, i.e. there is no overlap between these values). Analogously, they contrast with respect to the feature PaBr (all Australian languages have the value A while all the Daic languages have L) and CrCo (for similar reasons). The two language families do *not* contrast with respect to GrPa since their values for that feature overlap (the Daic Siamese has the same value, C, which occurs in all Australian languages), and, analogously, with respect to Sibl: the feature value Sibl=A in the Daic language Lao, recurs in the Australian languages Groote Eylandt, Tiwi and Kariera.

As far as contrasts involving features which have *unkown* / *missing* values are concerned (for example, those occurring in the Daic family languages Lao and Siamese for the feature Sb-II), MPD adopts a conservative strategy of matching and ignores a missing value, in the sense that it will not contrast it with any other value, including a missing value in the other compared family,

despite the fact that a contrast could in principle exist if the pattern that is at present unknown is subsequently acquired and is eventually found to match. This conservative approach is preferred, since it guarantees a higher reliability of profiles. At the same time, MPD uses the reliable, known information about patterns, even if some patterns for a value are missing. Thus, Australian and Daic *will* contrast with respect to the feature SbCh, even if the value for Lao for that feature is unknown, since no two languages, one belonging to one family and one to the other family, share a value. The value for Lao for that feature, “?”, is ignored as unknown. The two language families, however, will not contrast with respect to the feature Sb-II, as already noted, since there is no available information for positing (or rejecting for that matter) a contrast, as both Daic languages have unknown values for this feature.

If some pairs of language families fail to be discriminated by primitive features alone, MPD will recourse to forming complex, or derived features, and will check whether the families can be discriminated by these invented features. E.g. the Australian family cannot be demarcated from the Austronesian family by any one of the above primitive features (for brevity, not all the feature-value descriptions of the Austronesian languages, which are 76 in number in the dataset, will be enumerated). In this situation, MPD will combine any two primitive features into a derived nominal feature. For example, one such derived feature is Sibl- $\times$ -Sb-II, which for the Australian families will have the values shown below, as can be verified from the respective values for Sibl and Sb-II above:

Table 2: The complex feature Sibl- $\times$ -Sb-II of Australian languages

<i>Australian</i>	<i>Sibl-<math>\times</math>-Sb-II</i>
Groote Eylandt	(A + E)
Tiwi	(A + E)
Aranda	(N + D)
Dieri	(D + E)
Kariera	(A + E)
Murinbata	(B + E)

As it happens, none of the 76 languages in the Austronesian family has a value overlapping with one of these, so the two families are discriminated by this derived feature. It is to be noted that the values of derived features, just as those of primitive ones, may be unknown and contain “?”; the treatment of missing values here is the same as with primitive features, i.e. “?” serves neither to posit nor to reject a contrast.

## 6 Analysis of results

Using the MPD program, a large-scale computational discrimination of the language families in terms of their kin term patterns, as described in Murdock’s database, was conducted. Below, the results are briefly discussed with respect to the three questions posed in Section 3 (a detailed description of the profiles of families can be found in Pericliev, 2011).

### 6.1 The potential of kin term patterns for language family discrimination

64 language families with 494 attendant languages were profiled, using both primitive and complex features. Disposing of 64 classes to discriminate, 2016 pair-wise contrasts had to be made (the number of contrasts among  $N$  classes can be calculated by the formula  $N(N - 1)/2$ , or

$64(64 - 1)/2 = 2016$ ). The output of MPD shows that only 23 pairs of families remain undifferentiated, a result which strongly suggests that kin term patterns are a fairly good discriminant of language families. The non-discriminated pairs are the following:

(1) Afro-Asiatic and Indo-European, (2) Afro-Asiatic and Niger-Congo, (3) Anda-manese and Austronesian, (4) Arawakan and Khoisan, (5) Arawakan and Nambiquaran, (6) Arawakan and Salishan, (7) Austronesian and Choco, (8) Austronesian and Huaven, (9) Austronesian and Indo-European, (10) Austronesian and Niger-Congo, (11) Austronesian and Nilo-Saharan, (12) Austronesian and Sino-Tibetan, (13) Carib and Siouan, (14) Indo-European and North-Caucasian, (15) Indo-European and Niger-Congo, (16) Indo-European and Siouan, (17) Kiowa-Tanoan and Na-Dene, (18) Na-Dene and Niger-Congo, (19) Nambiquaran and Siouan, (20) Niger-Congo and Nilo-Saharan, (21) Niger-Congo and Sino-Tibetan, (22) Niger-Congo and Trans-New-Guinea, (23) Niger-Congo and Uto-Aztecan.

The kin term pattern overlap for some pairs of families may be due to mere chance, especially in cases of families comprising many languages, or it may be due to linguistically more interesting reasons, borrowing or genetic affiliation, for example. This problem will not be discussed here, except to note that 13 of these pairs of families belong to the same geographical area (pairs 1–3, 9–12, 14–15, 20–21 are in the areal III. Africa, Europe and Asia, and Kiowa-Tanoan and Na-Dene are in the areal I. North and Central America). Also, without wanting to make any far-reaching implications for the diagnostic nature of kinship semantics, it might be interesting to note that for some of these families more distant genetic relations have been suggested in literature (cf. e.g. Birnbaum, 1978; Bomhard, 1984 for Afro-Asiatic and Indo-European; Schuhmacher, 1974 for Austronesian and Indo-European; Boyd, 1978; Gregersen, 1972; Homburger, 1941 for Niger-Congo and Nilo-Saharan, etc.).

## 6.2 The contribution of features in family discrimination

The minimum number of overall features needed for discrimination is 20 (all 8 primitive features are used, as well as 12 complex, or derived (nominal) ones). These are:

1. Cross-Cousin
2. Grandchild
3. Grandparent
4. Nephew/Niece (male speaker)
5. Sibling-in-law
6. Uncle
7. Aunt
8. Sibling
9. Cross-Cousin- $\times$ -Sibling-in-law
10. Grandchild- $\times$ -Sibling-in-law
11. Grandchild- $\times$ -Nephew/Niece (male speaker)
12. Grandchild- $\times$ -Sibling
13. Grandparent- $\times$ -Sibling
14. Uncle- $\times$ -Sibling-in-law







	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Trans-New-Guinea	+		+	+	+			+										+		+
Tucanoan	+			+				+	+											
Tupi	+	+		+	+	+		+				+								+
Uralic	+			+		+						+								
Uru-Chipaya	+			+		+		+		+										
Uto-Aztecan	+			+	+		+	+		+		+		+					+	
Wakashan				+	+	+		+			+				+					
Witotoan				+																
Yanomam				+	+			+											+	
Yukaghir				+		+														
Yuki				+	+			+		+										
TOTAL	30	17	10	54	42	29	12	47	6	21	5	10	7	14	+17	10	5	25	7	26

### 6.3 The idiosyncrasy of language families

The MPD computer program guarantees that the *minimum* number of distinctive features will be used to individualize every family with a profile. Therefore, the number of distinctive features in a profile serving to demarcate a language family can be used as an *index of uniqueness*, measuring the idiosyncrasy, typicality, or the ease with which a family is differentiated from all others: the higher the index of uniqueness, the smaller the number of distinctive features a family has in its profile. The lower bound, giving the highest index of uniqueness, will of course be 1, as one is the minimum number of distinctive features that might be used in a profile. The upper bound will be equal to the largest number of distinctive features used in a profile.

Table 3 allows one to compute the index of uniqueness of each profiled family, by counting the number of distinctive features used in its profile, or row, in the table. According to this index, all language families can be ordered according to their “typicality” regarding their kin term patterns. In Table 4, the families are given in descending order of their index of uniqueness.

The index of uniqueness in the studied domain varies, ranging from 1 feature (Keres) to 14 features (Niger-Congo). The language family Keres is represented in Murdock’s database by a single language, Cochiti, and this language possesses a unique pattern for the primitive feature of grandparents, one distinguishing between grandparents on the basis of the speaker’s sex, grandfather (woman speaking) and grandmother (man speaking). This sets Keres apart from all the remaining families. The Niger-Congo family is represented in the database by 84 languages, and its 14 distinctive features include both primitive (5 in number) and derived features (9 in number). For all of the language families, the average number of distinctive features needed to discriminate any one of them is approximately 6 features per family.

The number of features in a profile, obviously, will depend to a certain extent on the size of the language family discriminated: the larger the family (i.e. the more languages it has as members), the more features will be needed for demarcation, and the smaller the family, the fewer the features needed. Thus, Niger-Congo, the biggest family with 84 members, has the lowest index of uniqueness and requires 14 features, while Keres, with only 1 language, has the highest index and requires 1 feature. However, this is apparently not the whole story. A large family like Austronesian, with 76 languages, requires just 9 features (in contrast to Niger-Congo’s 14), and the small Mixe-Zoque family, like Keres with 1 language, requires as many as 7 features (in contrast to Keres’s 1). It is therefore necessary to conclude that the index of uniqueness does not merely reflect the size (and time-depth) of a family, but also reflects the tendency within a family to develop historically in its own, idiosyncratic way.

Table 4: The index of uniqueness of language families

<i>Index of uniqueness</i>	<i>Language family</i>
1	Keres
2	Aymaran Chukotko-Kamchatkan Coahuiltecan Totonacan Witotoan Yuk- aghir
3	Araucanian Choco
4	Basque Caddoan East-Papuan Hmong-Mien Huavean Jivaroan South- Caucasian Tucanoan Uralic Yanomam Yuki
5	Australian Barbacoan-Paezan Chon Daic Dravidian Gulf Iroquoian Mus- kogeon Sepik-Ramu Uru-Chipaya
6	Andamanese Macro-Ge Nambiquaran Oto-Manguean Salivan Wakashan
7	Arawakan Mataco-Guaicuru Mixe-Zoque North-Caucasian Nilo-Saharan Salishan Trans-New-Guinea
8	Altaic Austro-Asiatic Eskimo-Aleut Hokan Japanese Kiowa-Tanoan Na- Dene Penutian Sino-Tibetan Siouan Tupi
9	Austronesian Carib Chibchan Indo-European Khoisan Uto-Aztecan
10	Afro-Asiatic Algic Mayan
14	Niger-Congo

## 7 Conclusions

The fundamental findings of this paper can be summarized as follows: (1) language families can be discriminated on the basis of their kin term patterns reasonably effectively, and therefore these patterns may be considered to be strong indicators of genetic affiliation, as suggested previously by the anthropologist G. P. Murdock. (2) The eight atomic and twelve complex features, twenty features in all, are the smallest number of features needed to discriminate all the language families. (3) The size of the profiles, comprising of exactly the necessary and sufficient conditions for demarcation, allows for the creation of “an index of uniqueness for a family”, ordering the language families according to the ease of their differentiation from the rest of the families. (4) The kin term patterns for the features “Nephews and Nieces (=siblings’ children)”, “Siblings”, and “Siblings-in-law” — i.e. all which include the idea of siblings — are the best predictors of genetic affiliation, as they are significantly more frequently used in the profiles, and all invented twelve complex features needed for discrimination must involve a combination of these features with each other, or with other features.

## References

- Birnbaum, H. (1978). *Linguistic reconstruction: Its potentials and limitations in new perspective*. Washington, D.C: Journal of Indo-European Studies Monograph, No. 2.
- Bomhard, A. R. (1984). *Toward Proto-Nostratic*. Amsterdam: John Benjamins.
- Boyd, R. (1978). *Etudes comparatives: A propos des ressemblances lexicales entre langues Niger-Congo et Nilo-Sahariennes*. Paris: SELAF. (Bibliothèque de la SELAF, 65).
- Gregersen, E. A. (1972). Kongo-Saharan. *Journal of African Languages*, 11(1), 69–89.
- Homburger, L. (1941). *Les langues negro-africaines*. Paris: Payot.
- Kirchhoff, P. (1932). Verwandtschaftsbezeichnungen und Verwandtenheirat. *Zeitschrift für Ethnologie*, 64, 41–72.

- Lowie, R. H. (1928). A note on relationship terminologies. *American Anthropologist*, 30(2), 256–66. <https://doi.org/10.1525/aa.1928.30.2.02a00060>
- Marshall, M. (1984). Structural patterns of sibling classification in island Oceania: Implications for culture history. *Current Anthropology*, 25(5), 597–637. <https://doi.org/10.1086/203199>
- Morgan, L. H. (1870). Systems of consanguinity and affinity of the human family. *Smithsonian Contributions to Knowledge*, 17, 1–590.
- Murdock, G. P. (1949). *Social structure*. New York: The MacMillan Company.
- Murdock, G. P. (1968). Patterns of sibling terminology. *Ethnology*, 7(1), 1–24. <https://doi.org/10.2307/3772805>
- Murdock, G. P. (1970). Kin term patterns and their distribution. *Ethnology*, 9(2), 165–208. <https://doi.org/10.2307/3772782>
- Pericliev, V. (2010). *Machine-aided linguistic discovery: An introduction and some examples*. London: Equinox.
- Pericliev, V. (2011). *Profiling language families by their kin term patterns: A computational approach*. Muenchen: Lincom EUROPA. (LINCOM Etymological Studies (LES), 02).
- Schuhmacher, W. (1974). B ~ C? (A=Indoeuropean, B=Austronesian, C=Eskimo). *Anthropos*, 69, 625–627.
- Spier, L. (1925). The distribution of kinship systems in North America. *University of Washington Publications in Anthropology*, 1(2), 69–88.
- Valdes-Perez, R. E., Pereira, F., & Pericliev, V. (2000). Concise, intelligible, and approximate profiling of multiple classes. *International Journal of Human-Computer Studies*, 53(3), 411–436. <https://doi.org/10.1006/ijhc.2000.0397>

---

## Acknowledgment

This work was supported by core funding for statutory activities from the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences.

The author declares that he has no competing interests.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (<http://creativecommons.org/licenses/by/3.0/pl/>), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

© The Author 2017