Serdica Journal of Computing 15(1), 2021, pp. 27-42, 10.55630/sjc.2021.15.27-42
Published by Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

Machine Componential Analysis of Kinship Vocabularies: An Example

Vladimir Pericliev

Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences, Bulgaria
peri@math.bas.bg

Abstract

The paper introduces a computer program to handle the task of componential analysis of kinship terms. Given the kin terms of a language with their attendant kin types, the program discovers all componential models, including the "simplest" ones (using the minimum number of dimensions and components in kin term definitions). As an illustrative example of the application of the program we use the kinship vocabulary of Catalan, a previously unanalyzed language. A completely unconstrained analysis of Catalan leads to an intolerably large number of alternative models but our simplicity restriction pertaining to choosing the minimum number of dimensions leads to a unique model. The generated componential model of Catalan uses four dimensions of contrast, viz. "sex", "generation", "distance" and "affinity", and coincides with that of Spanish. However, it is different from those of other Indo-European languages that also use for demarcation this very set of dimensions.

Keywords: Machine Scientific Discovery, Machine Learning

ACM 2012 CCS Concepts: Computing methodologies → Artificial intelligence → Natural language processing

Mathematics Subject Classification 2020: 91F20, 03B65, 91D99, 68T01, 68T20

Received: May 3, 2022, *Accepted:* May 10, 2022, *Published:* May 17, 2022

Citation: Vladimir Pericliev, Machine Componential Analysis of Kinship Vocabularies: An Example, Serdica Journal of Computing 15(1), 2021, pp. 27-42, https://doi.org/10.55630/sjc.2021.15.27-42

1 Introduction

A method of structural semantics, known as "componential analysis", is based on the Saussurean idea of "system" in linguistics, in which the objects forming a system are described as a conjunction of smaller components that are necessary and sufficient to distinguish each object in the system from all others. Kinship terminological studies from a structural perspective were initiated by the classical works of Goodenough [1, 2], Lounsbury [3] and some others and have gained wide popularity in both linguistics and anthropology, resulting in a significant number of componential models proposed in the literature over the years for mostly "exotic" languages.

Some analysts believe that these models reveal "psychological validity", or the world view of native speakers, while others assume that such models describe "social-structural reality", or simply the rules of using kin terms in a society, and the discovery of psychologically valid models requires subsequent psychological tests. Some scholars have proposed such tests, while others have introduced alternative ways to study kinship terminology (extensionist, algebraic or relational). Despite the existence of alternative approaches, classical structural (componential) analysis is indispensable to kinship studies and is practiced by linguists and anthropologists of different theoretical persuasions. There are several reasons for this.

First, componential analysis is an inalienable part of some of the other approaches (for instance the extensionist method presupposes componential models of the "core vocabulary" while the approaches looking for psychological validity presuppose the availability of componential models to be subsequently tested psychologically). Secondly, it is the only method that reveals the semantic system of kin terms and for this reason continues to be in the analytic repertoire of both current linguistics and current anthropology. And, thirdly, componential models are essential not only for exposing semantic structure, but also for translation purposes and for historical semantic reconstruction.

The goal of the present paper is to briefly present a computer program that automates the task of componential analysis and to analyze Catalan kinship terminology with this program. A similar discussion, using Serbo-Croatian as an example, can be found in [4]. Catalan is a western Romance language, spoken by some 4.1 million people in Spain, Andorra, France and Italy. The discovery of the structure of kinship systems has been recognized as a difficult problem for human analysts and Leech [5, p. 239] e.g. writes that "kinship analyses have a mind-teasing quality of mathematical puzzles. The only cure for bafflement is to think hard and hope that the light will dawn!".

The paper is organized as follows. Section 2 introduces the basic notions, Section 3 sketches the computer program for componential analysis, and Section 4 presents the analysis generated by the program. Finally, in Section 5, we briefly compare the obtained analysis of Catalan with those of some other languages explored with the same computational means.

2 The basic notions of componential analysis

Componential analysis of kinship terminology is a well-known method familiar from the works of Goodenough [2, 6], Lounsbury [3], Wallace and Atkins [7], Hammel [8], Leech [5], and more recently, Geeraerts [9] and Bernard [10].

The "kin terms" of a language, such as Catalan mare 'mother', pare 'father', oncle 'uncle', etc., are linguistic labels for a range of "kin types", which specify the genealogical position of one's kin with respect to oneself. In the following, we use the standard abbreviations [11] of atomic genealogical relationships in terms of which the kin types are expressed:

Fa = 'father', Mo = 'mother', Br = 'brother', Si = 'sister',

So = 'son', Da = 'daughter', Hu = 'husband', and Wi = 'wife'.

These atomic relationships are juxtaposed to express more distant kin types (relatives), as, e.g., MoBr 'mother's brother', MoSi 'mother's sister', MoSiHu 'mother's sister's husband', etc.

The meaning of kin terms is represented by all kin types, or relatives, covered by the term. For example, the meanings of the Catalan kin term *oncle* is MoBr or FaBr. The set of all kin terms in a language is the "kinship vocabulary" of the language.

The *basic goal* of componential analysis is to determine the relevant conditions for distinguishing the meaning of any of the kin terms within the kinship vocabulary from any other. Put differently, componential analysis should find, for any kin term, the common features for all its attendant kin types, such that these common features demarcate this term from all other kin terms in the kinship vocabulary. From such a componential model, given any pair of individuals in a society, alongside with some (minimum) items of information about their relationship, such as their sex, generation, etc., the analyst should be able to infer whether they are kinfolk and what terms they would use to refer to one another.

As in most other grammatical tasks, a general common *adequacy requirement* to componential analysis would be to discover, for any data set, all and only the componential paradigms that describe the structure of the domain. This means that proposed componential models should be "consistent", i.e. kin terms must be defined by necessary and sufficient features, and besides all componential models for a given data set should be revealed.

Additionally, as in other grammatical tasks, proposed componential models should be the simplest, as simplicity is a highly evaluated virtue of linguistic analyses. The requirement for parsimony in our case embodies two basic criteria: (1) choose the smallest number of overall contrasting features (dimensions) sufficient to describe the kinship vocabulary; and (2) choose the smallest number of components in the definitions of every kin term in the vocabulary.

3 The computer program

3.1 An overview of the program

We have implemented a computer system, whose aim is to attempt to resolve both the problem of consistency of componential models and to generate all alternative models, eventually trying to resolve the multiplicity of solutions problem.

The system accepts as input the set of kin terms for a language with all their attendant kin types. Given this information, the system can generate all consistent componential models for the set, applying natural simplicity criteria to constrain the choices in case of alternatives. The system is actually an extension of a sophisticated general-purpose class discrimination program that has found various applications in linguistics and outside of linguistics. Below we confine to a brief description of the program, computational details being discussed elsewhere [12, 13, 14].

The program is endowed with a set of features (or dimensions) and with subroutines that determine, for each kin type, the value the kin type has for the inspected feature. For instance, the values of a kin type for the feature "sex" can be determined by the system by its last symbol (=link), knowing further the sex of all atomic relationships. Thus, the program can find that the kin type FaBr is sex=male, since its last link, viz. Br, is male, while FaSi is sex=female, since Si is female. The feature "generation" of a kin type is determined as a sum of the generations of the links constituting this kin type, where the latter are $+1$ for the parental relationships Fa and Mo, -1 for the filial relationships So and Da, and 0 for all remaining relationships Br, Si, Hu, and Wi; thus the program can compute that the kin type FaBr is generation=1, since $+1+0=1$. Similarly, all feature values (components) are computed, and this is done for all kin types in the data set.

The program then transfers the components of the kin types into components of kin terms by finding those components that are possessed by all the kin types covered by a kin term. For example, a Catalan kin term like oncle will have the components [sex = male & generation = 1 & distance = 2 & affinity = consanguineal], since both its kin types, viz. FaBr and MoBr, are both male sex, one generations above ego, genealogical distance of two and consanguineal (blood relatives).

The program then proceeds with computing the dimensions of contrast, or contrasting features, that demarcate each kin term from each other term in the dataset. Having found these pair-wise contrasts between kin terms, the definition of each term is produced, which contains the necessary and jointly sufficient contrasting features (or components) that discriminate this term from all others in the dataset. The computation is complex because it involves the finding of a minimum set cover, which is known to be an NP-complete problem. In the worst-case, it is computationally intractable, but the cases with kinship vocabularies we consider are easy to handle with our program.

3.2 The features used by the program

Currently, our program employs fifteen features. It uses those of Kroeber [15], viz., "generation", "lineal versus collateral", "age difference in one generation", "sex of the relative", "sex of the first connecting relative", "sex of the speaker", "consanguineal versus affinal". Greenberg [16, p. 13] writes about these features: "Leaving aside some difficulties and complications, in principle any kin term in any language can be specified by means of them". To alleviate some potential difficulties envisaged by Greenberg, we have also included the features "distance", "sex of the second connecting relative", and some others in order to handle languages of various types (Eskimo, Sudanese, Hawaiian, etc.). Below we list only those features that are used for the purposes of this paper. Some of these are self-explanatory, whereas others are explained by simple examples.

(1) Generation of relative, with feature values

generation = 1

generation = 0

generation = −1, etc.

The value of the feature "generation" can be any integer, including a range of integers (bounded by $\geq$ 'equal or greater than' or $\leq$ 'smaller or equal to') to handle cross-generational kin terms.

(2) Sex of relative, with feature values

sex = m – male

sex = f – female

(3) Genealogical distance, with the integer feature values

distance = 1

distance = 2, etc.

This feature, analogously to "generation", can take as value any integer, including a range of integers (bounded by $\geq$ 'equal or greater than' or $\leq$ 'smaller or equal to'). The kin type Fa has the value distance = 1, FaBr has the value distance = 2, and FaBrWi distance = 3.

(4) Affinity of relative, with two feature values

affinity = cons – consanguineal, or absence of a marital tie

affinity = aff – affinal, marital tie

Examples: the kin type MoSi has the value affinity = cons, while Wi and SiHu have the feature value affinity = aff.

(5) Affinity of the 1st connecting relative (link), with feature values

affinity 1st link = cons – consanguineal (first link is a blood relative)

affinity 1st link = aff – affinal (first link is a relative by marriage)

Examples: the kin types <u>Fa</u>Si, <u>Si</u>Hu (where first link is underlined) both have the feature value affinity 1st link = cons, while <u>Wi</u>Mo, <u>Hu</u>Si are both affinity 1st link = aff.

(6) Sex of the 1st connecting relative (link), with feature values

sex 1st link = m – male

sex 1st link = f – female

Examples: the kin types <u>Fa</u> and <u>Fa</u>Si (where first link is underlined) both have the feature value sex 1st link = m, while <u>Mo</u> and <u>Wi</u>Si are both sex 1st link = f.

(7) Generation of the last link, with the feature values

generation last link = 1

generation last link = 0

generation last link = -1

The generations of the last links will be: Fa and Mo = 1; Si, Br, Hu and Wi = 0; So and Da = −1; hence, e.g., Fa<u>Mo</u> (where last link is underlined) will have the feature value generation last link = 1 for this feature, while <u>So</u> or So<u>So</u> will have the feature value generation last link = −1.

(8) Sex of the second connecting relative (link), with feature values

sex 2nd link = m – male

sex 2nd link = f – female

Examples: the kin types Fa<u>Fa</u>Si, Mo<u>Br</u> (where second link is underlined) both have the feature value sex 2nd link = m, while Fa<u>Mo</u>Si, Mo<u>Si</u> are both sex 2nd link = f.

(9) Affinity of the last connecting relative (link), with feature values

affinity last link = cons – consanguineal

affinity last link = aff – affinal

Examples: the kin types FaFa<u>Br</u>, Mo<u>Br</u> (where last link is underlined) both have the feature value affinity last link = cons, while FaBr<u>Wi</u>, Si<u>Hu</u> are both affinity last link = aff.

(10) Lineality, with the feature values

lineality = lin – lineal

lineality = coll – collateral

3.3 The simplicity constraints of the program

Our system uses three intuitive criteria to guarantee the uncovering of the simplest discrimination of the kinship terms. They refer to dimensions (first criterion) and components in kin term definitions (second and third criterion)

1. *Minimize overall features (=dimensions).* A set of kin terms may be demarcatable, using a number of overall feature sets of different cardinality; this criterion chooses those overall feature sets which have the smallest cardinality (i.e. are the shortest).

2. *Minimize features (=components) in kin term definitions.* Given some overall feature set, one kin term may be demarcatable – using only features from this set – by a number of definitions of different cardinality; this criterion chooses those definitions, having the smallest cardinality (i.e. are the shortest).

3. *Minimize "minor" features in kin term definitions.* This criterion minimizes the use of "minor", i.e. infrequent, peripheral, features in kin term definitions, in the case when alternatives remain even after the application of the previous three simplicity criteria. The basic idea is that in the alternative, equally short definitions, of one kin term, some features are obligatory and must be used, while others are optional and may or may not be used. From the optional features for one kin term, the criterion prefers the more frequent feature/s from a frequency hierarchy computed from the obligatory features used in the whole componential scheme (see below).

3.4 The program as a computational tool

Our program has a number of facilities that make it a convenient research computational tool. Importantly, the program includes a mechanism for inventing derived features by combining old features by means of the logical operations conjunction, disjunction, implication, and equivalence. This mechanism, reflecting interactions between features, can be used when the available features are insufficient to discriminate all kin terms Also, the user can present a number of queries to the program, such as:

- What are the attendant kin types of a given kin term, or what kin term corresponds to a given set of kin types?
- Does cross-classification of kintypes occur, i.e., is one kin type classed under two or more different kin terms? (Such data precludes complete discrimination of kin terms, so the system displays the faulty kin types.)
- What are the current features of the system?
- What kin type or kin term possesses specific features?
- What semantic contrasts exist between a selected pair of kin terms?
- What pairs of kin terms are indiscriminable?

Besides making these queries, the user is also free to employ a prespecified subset of the set of available features that is for some reason-structural, cultural, or psychological-considered important by the analyst. (This would allow an analyst to discover, for example, how the features suggested by Kroeber (1909) [15] fare when applied to a specific group of languages.)

4 The Catalan kinship terms and their analysis

The Catalan data subjected to componential analysis is given below. The dataset is pretty comprehensive, but probably not exhaustive (as noted by a referee of the paper). The kin terms of the language are listed with all their attendant kin types:

1. *besavi*: FaFaFa FaMoFa MoFaFa MoMoFa
2. *besàvia*: FaFaMo FaMoMo MoFaMo MoMoMo
3. *avi*: FaFa MoFa
4. *àvia*: FaMo MoMo
5. *besoncle*: MoMoBr MoFaBr FaMoBr FaFaBr
6. *bestia*: MoMoSi MoFaSi FaMoSi FaFaSi
7. *oncle*: MoBr FaBr
8. *tia*: MoSi FaSi
9. *mare*: Mo
10. *pare*: Fa
11. *germana*: Si
12. *germà*: Br
13. *cosina*: MoSiDa MoBrDa FaSiDa FaBrDa MoMoSiDaDa MoMoSiSoDa MoMoBrDaDa MoMoBrSoDa MoFaSiDaDa MoFaSiSoDa MoFaBrDaDa MoFaBrSoDa FaMoSiDaDa FaMoSiSoDa FaMoBrDaDa FaMoBrSoDa FaFaSiDaDa FaFaSiSoDa FaFaBrDaDa FaFaBrSoDa
14. *cosi*: MoSiSo MoBrSo FaSiSo FaBrSo MoMoSiDaSo MoMoSiSoSo MoMoBrDaSo MoMoBrSoSo MoFaSiDaSo MoFaSiSoSo MoFaBrDaSo MoFaBrSoSo FaMoSiDaSo FaMoSiSoSo FaMoBrDaSo FaMoBrSoSo FaFaSiDaSo FaFaSiSoSo FaFaBrDaSo FaFaBrSoSo
15. *fill*: So
16. *filla*: Da
17. *nebot*: BrSo SiSo
18. *neboda*: BrDa SiDa
19. *nét*: SoSo DaSo
20. *néta*: SoDa DaDa
21. *besnét*: SoSoSo SoDaSo DaSoSo DaDaSo
22. *besnéta*: SoSoDa SoDaDa DaSoDa DaDaDa

23. *marit*: Hu
24. *esposa*: Wi
25. *sogra*: WiMo HuMo
26. *sogre*: WiFa HuFa
27. *gendre*: DaHu
28. *nora*: SoW
29. *cunyat*: HuBr WiBr
30. *cunyada*: HuSi WiSi

Running our system on this dataset, with its 15 features and no simplicity constraints, reveals a massive indeterminacy insofar as overall features, or dimension sets, are concerned. Thus, there are 18 such sets, comprising necessary and jointly sufficient features to demarcate all Catalan kin terms. These are listed below:

A. {sex & distance & generation & affinity}
B. {sex & generation last link & generation 1st link & distance & affinity 1st link}
C. {sex & generation 1st link & generation & distance & affinity 1st link}
D. {sex & generation last link & generation & distance & affinity 1st link}
E. {sex & generation last link & generation 1st link & distance & affinity last link}
F. {sex & affinity 1st link & generation & distance & affinity last link}
G. {sex & generation 1st link & generation & distance & affinity last link}
H. {sex & generation last link & generation & distance & affinity last link}
I. {sex & generation last link & distance & generation 1st link & affinity}
J. {sex & generation last link & generation 1st link & generation & affinity 1st link & affinity last link}
K. {sex & generation 1st link & lineality & generation & affinity 1st link & affinity last link}
L. {sex & generation last link & lineality & generation & affinity 1st link & affinity last link}
M. {sex & generation last link & lineality & affinity 1st link & distance & affinity last link}

N. {sex & generation last link & lineality & distance & affinity 1st link & affinity}

O. {sex & generation last link & lineality & distance & affinity last link & affinity

P. {sex & generation last link & generation 1st link & affinity last link & generation & affinity}

Q. {sex & generation 1st link & lineality & affinity last link & generation & affinity}

R. {sex & generation last link & lineality & affinity last link & generation & affinity}

To each overall feature (dimension) set, there correspond distinct alternative componential models, resulting from the different definitions that can be given to some kin terms, using only features from this overall feature set. In one componential model, each distinct definition of a kin term may freely combine with any other alternative definition of all other terms. Thus, the total number of the alternative componential models, Q, using just one dimension set, would be equal to the product of the number of definitions, N, each individual term has obtained, expressed by the formula $Q = N_1 \times N_2 \times N_3 \cdots \times N_m$.

For example, assuming our dataset to comprise only three terms, the first having two definitions, the second one, and the third four, we have $Q = 2 \times 1 \times 4 = 8$ componential analyses in all. In our particular case, if we consider just the alternative componential models corresponding to dimension set B above, we have 13 kin terms with 2 definitions and 2 terms with 3 definitions (for brevity, not given here), i.e. $2^{13} \times 3^2 = 73728$ models. This is a large number of alternative models and there would be still many others corresponding to the other dimension sets.

This result is in accordance with the warning as far back as Burling [17] against the multiplicity of solutions problem. In a paper entitled "Cognition and componential analysis: God's truth or hocus pocus?", Burling tried to show the large number of logically possible alternative componential models of any given set of kin terms. Thus, if there are three items in the kin term set (call the items a, b and c), one has three apparent choices: use a component which separates a from b and c; one which separates b from a and c; or one which separates c from a and b. The possibility of using components which are relevant for only a part of the set doubles the number of possibilities.

Whereas for a three-term set the possibilities number 6, for any given four kin term set they number 124. In general, the number of logically possible alternative componential models steeply increases with the increase of the number of kin terms that have to be discriminated. In Burling's opinion, there are no means to sensibly reduce this huge number of alternatives and the methods advocated are not equal to their goal. The conduct of componential analysis is therefore "hocus-pocus" rather than an enterprise that reveals "God's truth".

However, our three simplicity constraints outlined above generally resolve the multiplicity of solutions problem as we have observed applying our computer program to many languages of diverse structural types. In the case of Catalan, our simplicity criterion pertaining to overall feature sets (constraint 1) chooses the shortest set, viz. Set A, comprising four features (the others being with five and six component members). Using this four-member set, all kin terms obtain unique definitions, with no need to apply a further simplicity constraint. The resultant unique simplest componential analysis is stated in Table 1.

The analysis on Table 1 reveals the interrelationships, or structure, of the field of Catalan kinship terms. Four features (dimensions), viz. "sex", "generation", "genealogical distance" and "affinity", are necessary and sufficient to discriminate all 30 kin terms. Each of these terms is defined contrastively, with features that are necessary and sufficient to demarcate the term from all remaining terms. Thus, for instance, for the term *nét*, defined as [sex = m & generation = -2], the two components are jointly sufficient for singling it out, since the feature generation = -2 demarcates *nét* from all terms that are not generation = -2, which are all terms except *néta*, and sex = m distinguishes it from *néta*, which is female. Each of the two components in the definition of *nét* are also necessary, since leaving out sex = m will fail the discrimination with *néta*, while leaving out generation = -2 will fail the discrimination with all terms with the feature value sex = m (*besavi*, *avi*, etc.).

The four dimensions our program has chosen in the simplest model are well justified. "Sex", "generation", "genealogical distance" and "affinity" symmetrically and economically partition the space of relatives in Catalan and, additionally, are common cross-linguistically. The alternative dimension sets are merely versions of the basic, simplest, set, as they comprise variants of these dimensions. For instance, the features "generation 1st link" and "generation last link" – though absolutely necessary for describing other languages – capture the general idea of generation, which the feature "generation" already does, hence are not required for understanding the componential structure of Catalan. Ignoring dimension sets with such "subsidiary" features, filters out all dimension sets except Set A. Importantly, Set A is the simplest one, chosen by our first simplicity criterion.

Table 1: Simplest componential analysis of Catalan. Abbreviations:
m = male, f = female, cons = consanguineal (blood), aff = affinal, & = and.

1.	*besavi*	sex = m & generation = 3
2.	*besàvia*	sex = f & generation = 3
3.	*avi*	sex = m & generation = 2 & distance = 2
4.	*àvia*	sex = f & generation = 2 & distance = 2
5.	*besoncle*	sex = m & generation = 2 & distance = 3
6.	*bestia*	sex = f & generation = 2 & distance = 3
7.	*oncle*	sex = m & generation = 1 & distance = 2 & affinity = cons
8.	*tia*	sex = f & generation = 1 & distance = 2 & affinity = cons
9.	*mare*	sex = f & generation = 1 & distance = 1
10.	*pare*	sex = m & generation = 1 & distance = 1
11.	*germana*	sex = f & generation = 0 & distance = 1 & affinity = cons
12.	*germà*	sex = m & generation = 0 & distance = 1 & affinity = cons
13.	*cosina*	sex = f & generation = 0 & distance $\geq$ 3
14.	*cosi*	sex = m & generation = 0 & distance $\geq$ 3
15.	*fill*	sex = m & generation = -1 & distance = 1
16.	*filla*	sex = f & generation = -1 & distance = 1
17.	*nebot*	sex = m & generation = -1 & distance = 2 & affinity = cons
18.	*neboda*	sex = f & generation = -1 & distance = 2 & affinity = cons
19.	*nét*	sex = m & generation = -2
20.	*néta*	sex = f & generation = -2
21.	*besnét*	sex = m & generation = -3
22.	*besnéta*	sex = f & generation = -3
23.	*marit*	sex = m & distance = 1 & affinity = aff
24.	*esposa*	sex = f & distance = 1 & affinity = aff
25.	*sogra*	sex = f & generation = 1 & affinity = aff
26.	*sogre*	sex = m & generation = 1 & affinity = aff
27.	*gendre*	sex = m & generation = -1 & affinity = aff
28.	*nora*	sex = f & generation = -1 & affinity = aff
29.	*cunyat*	sex = m & generation = 0 & distance = 2
30.	*cunyada*	sex = f & generation = 0 & distance = 2

5 A comparison of Catalan with other languages

We may briefly compare the componential model of Catalan with the models of some other Indo-European languages we have analyzed with the same set of 15 features (cf. [14]). The compared languages are: English (West Germanic), Swedish (East Germanic), Irish (Celtic), Spanish (Italic), Polish and Czech (West Slavic), Bulgarian and Croatian (South Slavic), Persian (Indo-Iranian), Albanian and Armenian.

Catalan employs 4 dimensions of contrast, and there is only one language of the analyzed set that uses a smaller number of contrastive features (Irish uses 3 features: "sex", "generation", "distance"). Spanish, Czech, Persian and Albanian achieve the discrimination by means of 4 features as Catalan, while Armenian employs 5, Bulgarian and Polish 6 and English 7 features. As could be expected for members of one family, the languages generally use overlapping sets of dimensions. The dimensions not used by Catalan, but used by some of the other Indo-European languages, are "sex of 1st link" (Polish, Persian, Armenian, Bulgarian, Croatian), "sex of 2nd link" (Polish, Armenian, Croatian) and "generation of last link" (English, Bulgarian).

To give just one example, the dimension "sex of 2nd link" (along with the dimension "sex of 1st link") is necessary in Polish to discriminate the lexemes for male cousin, viz. *brat stryjeczny* 'FaBrSo' (2nd link underlined), *brat cioteczny* 'FaSiSo or MoSiSo', and *brat wujeczny* 'MoBrSo', as well as those for female cousin, *sestra stryjeczna* 'FaBrDa', *sestra cioteczna* 'FaSiDa or MoSiDa', and *sestra wujeczna* 'MoBrDa'. (This feature is not popular worldwide, but would be necessary in a further Indo-European language like Hindi and in a non-Indo-European language like Turkish.)

Finally, we note that from the languages that share the same 4 contrastive dimensions with Catalan Spanish, its closest relative, has the same number of kin terms and the same semantic structure of its kinship lexical field (i.e. their kin terms receive identical componential definitions). The other languages with four-sized dimension sets have a different number of kin terms and different semantic structures than that of Catalan (and Spanish). In effect, some terms that are "translational equivalents" between these languages may turn out to be described with different sets of semantic components. For instance, the Catalan term *germà* 'Br' (brother) has the componential structure [sex = m & generation = 0 & distance = 1 & affinity = cons], while its translation in Czech *bratr* 'Br' has the structure [sex = m & generation = 0 & affinity = cons], lacking the feature "distance", because this feature is redundant in discriminating *bratr* from the remaining terms in the system of Czech kinship vocabulary.

6 Conclusion

We briefly introduced a computer program to handle the task of componential analysis of kinship terms and analyzed Catalan with this program. A completely unconstrained analysis of Catalan leads to an intolerably large number of alternative models in accordance with the warning of Burling (1964) [17] of the possibility of a huge number of alternatives. Our simplicity restriction pertaining to choosing the least number of dimensions, however, led to a unique model, thus rehabilitating the usefulness of the method of componential analysis. The generated structural model of Catalan uses four dimensions of contrast, viz. "sex", "generation", "distance" and "affinity", and coincides with that of Spanish. However, it is different from those of other Indo-European languages (Czech, Persian and Albanian) that also use for demarcation this very set of dimensions.

References

[1] W.H. Goodenough, "Componential Analysis and the Study of Meaning", *Language*, 32(1), 1956, pp. 195-216.

[2] W.H. Goodenough, "Componential Analysis of Könkämä Lapp Kinship Terminology", in W.H. Goodenough (ed.) *Explorations in Cultural Anthropology*, 1964, pp. 221-238, McGraw-Hill, New York.

[3] F.G. Lounsbury, "A Semantic Analysis of the Pawnee Kinship Usage", *Language*, 32(1), 1956, pp. 158-194.

[4] V. Pericliev, "Componential Analysis of Kin Terms – Some Problems and their Solutions", *Journal of Universal Language*, 15(1), 2014, pp. 129-168.

[5] G. Leech, *Semantics. England*, Pelican, Harmondsworth, 1974.

[6] W.H. Goodenough, "Componential Analysis: Kinship studies in cultural anthropology are producing a new tool for semantic analysis", *Science*, 156(3779), 1967, pp. 1203-1209.

[7] A.F.C. Wallace, J. Atkins, "The Meaning of Kinship Terms", *American Anthropologist*, 62(1), 1960, pp. 58-80.

[8] E.A. Hammel, *Formal semantic analysis*, American Anthropological Association, 1965.

[9] D. Geeraerts, *Theories of Lexical Semantics*, Oxford University Press, Oxford, 2009.

[10] H.R. Bernard, *Research methods in anthropology: Qualitative and quantitative approaches*, A division of Rowman & Littlefield Publishers, Oxford, 2011.

[11] G.P. Murdock, *Social Structure*, Macmillan, New York, 1949.

[12] V. Pericliev, R.E. Valdés-Pérez, "Automatic Componential Analysis of Kinship Semantics with a Proposed Structural Solution to the Problem of Multiple Models", *Anthropological Linguistics*, 40(2), 1998, pp. 272-317.

[13] R.E. Valdés-Pérez, F. Pereira, V. Pericliev, "Concise, Intelligible, and Approximate Profiling of Multiple Classes", *International Journal of Human-Computer Studies*, 53(3), 2000, pp. 411-436.

[14] V. Pericliev, *Componential Analysis of Kinship Terminology: A Computational Perspective*, Palgrave Macmillan, Basingstoke & New York, 2013.

[15] A.L. Kroeber, "Classificatory Systems of Relationship", *Journal of the Royal Anthropological Institute*, 39, 1909, pp. 77-84.

[16] J.H. Greenberg, "Universals of kinship terminology: Their nature and the problem of their explanation", in J. Maquet (ed.) *On Linguistic Anthropology: Essays in Honor of Harry Hoijer*, 1980, pp. 9-32, Udena Publications, Malibu.

[17] R. Burling, "Cognition and Componential Analysis: God's Truth or Hocus-Pocus?", *American Anthropologist*, 66(1), 1964, pp. 20-28.