In the previous chapter, I introduced the KINSHIP system and showed its successful operation on data from Bulgarian. In this chapter, I conduct componential analysis of a dozen other languages of Indo-European and non-Indo-European origin. The basic goals of these exercises in machine componential analysis are both to provide a further empirical test of the capacity of the program and to present ‘adequate’ models of the examined languages, many of which have not been analysed previously.

4.1 Introduction

As mentioned in Chapter 3, KINSHIP is endowed with about 20 overall features (dimensions). The following subset of 15 features was selected for the analysis of all languages below:

1. Sex of relative
2. Sex of speaker
3. Sex of 1st link
4. Sex of 2nd link
5. Generation of relative
6. Generation of 1st link
7. Generation of last link
8. Genealogical distance
9. Affinity of relative
10. Affinity of 1st link
11. Affinity of last link
12. Lineality
13. Seniority within one generation
14. Parallelity
15. Structural equivalence
Most of these features were explained in Chapter 3, but some further specifications may be added as required at appropriate places. Suffice it to say here that they include the standard features proposed by Kroeber (1909) (‘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’), but also some additional features pertaining to the first, second and last link of a kin type that turn out to be necessary in order to successfully demarcate the kin terms of specific languages.

The data subjected to study include the following languages: English (West Germanic), Swedish (East Germanic), Irish (Celtic), Spanish (Italic), Polish and Czech (West Slavic), Persian (Indo-Iranian), Albanian, Armenian (all belonging to the Indo-European), as well as Turkish (Altaic), Seneca (Iroquoian), Zapotec and Popoloca (Oto-Manguean), and Huave (language isolate). The set of examined languages includes predominantly Indo-European languages, covering to some extent most major branches of the family, and a couple of non-Indo-European languages. The choice of languages was partly determined by the fact that previous analysts have largely focussed on more ‘exotic’ languages, thus ignoring Indo-European ones, which, beyond doubt, also deserve attention, and partly because they cover the basic structural types (Eskimo, Sudanese, Iroquois, etc.). The particular data (kin terms and their attendant kin types) for the languages come from published accounts, dictionaries or other sources.

In the sequel, KINSHIP will start with the set of 15 features (dimensions) listed above and will output, for each language, the set of all alternative dimension sets that can serve for the discrimination of every kin term from its kinship vocabulary (thus highlighting its degree of ambiguity), as well as present its ‘simplest’ componential model(s), fulfilling our three simplicity criteria, pertaining to shortness of dimension set and kin term definitions and coherence of the latter. In case I am aware of previous componential analyses of the examined language, a concise comparison of the human and machine analyses will be made.

4.2 Exercises in machine componential analysis

4.2.1 English

English has been subjected to componential analyses in several works, notably by Wallace and Atkins (1960), Nogle (1974), Wordick (1973) and Goodenough (1965). The most comprehensive data set for American