

Conjoint analysis in linguistics – Multi-factorial analysis of Slavonic noun phrases

Tim Züwerink, M.A.

Bonn University

zuewerink@web.de

Decision processes between formally different but semantically comparable language structures are complex and influenced by a variety of factors. Some of these are associated with the speaker, others with the utterance or with the situation of communication. Though language variation is in primary focus of modern linguistics, comparably little effort has been made (for an exception see e.g. Gries 2001) on developing statistical procedures to reveal the importance of single grammatical, functional-stylistic or sociolinguistic factors in multi-factorial designs.

The conditions of alternation in particular Slavic languages between the adnominal genitive and the “possessive adjectives” (“PA”) have been characterised as very complex (see e.g. Corbett 1987 and Ivanov et al. (ed.) 1989). The aim of an online survey, conducted by the author with more than 1000 Russian (and additionally approximately 300 Czech and 300 Croatian native speakers), was a multi-factorial analysis of the decision process between the “PA” and the substantive in the genitive case without a preposition under varying lexical (“PA” and noun phrases), grammatical (different case, number and gender of noun phrases), functional-stylistical (different social spheres and registers of communication) and also sociolinguistic (respondents’ characteristics) circumstances.

Conjoint analysis (CA), which apparently has not been used in linguistics before, was applied as the survey technique. At first CA was applied in psychology (Luce and Tukey 1964). In market research, where it is most common and widespread, this type of analysis breaks down the customer’s decision between products, showing the importance and preference impact of different product attributes (see Backhaus et al. 2006). When confronted with a large number of decisions between varying product offers, the researcher will ideally be able to determine the importance of a particular factor in the customer’s eyes. The analysis is based on a regression, that presupposes, that every object (e.g. car) can be understood as the sum of its relevant factors (e.g. colour, brand, maximum speed and others), which results in the estimation of part-worth utilities for every factor value (e.g. “red”, “Mercedes”, “230 km/h”).

For the purpose of this analysis, it was assumed that language variation, just like choosing a product, is influenced by numerous factors, some of which can be varied systematically to show their relative importance and the preference impact of factor values for the speaker's decision between similar syntactic structures.

Every respondent was given a series of 30 stimuli after retrieving his personal data, which were evaluated as the respondents' characteristics. Respondents were chosen from their mother countries as well as Russians, Croats and Czechs, who spent a longer period of time abroad to estimate the importance of exposure to another language environment. The maximum possible independence of all influencing factors from each other was a crucial requirement. So "PA" had to be found, that represent the full spectrum of suffix types and semantical types of their base morphemes (these were proper names, kinship terms, other common nouns referring to persons, common nouns referring to animals).

More than 40000 responses were gathered in the three analyzed languages. On account of these data structures key factors of the usage of "possessive adjectives" with the suffixes *-ov-(-ův-)/-ev-* and *-in-/-yn-* in contemporary Russian, Croatian and Czech can be presented.

The paper will conclude with a description of possible applications for multi-factorial types of analysis to other problems in general linguistics. For example, speakers' preference for different word order patterns could possibly be analysed under varying circumstances. The factors that play a part in choosing between different classes of synonyms (for example, native and foreign words) could also be examined with the aid of this statistical approach. However, the semantics of the alternatives presented to the respondents, must be largely identical, which could bring difficulties in itself, given the fact that absolute synonymy is not possible. The researcher's task will be to indicate possible semantic differences and to keep them in a controllable area. In any case, this technique seems promising for revealing multi-factorial preference structures in any given language.

Backhaus, K. et al. (2006). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Elfte, überarbeitete Auflage, Berlin et al.

Corbett, G. (1987). The Morphology/Syntax interface: Evidence from Possessive Adjectives in Slavonic. *Language*, 63: 299-345.

Gries, S. (2001). A Multifactorial Analysis of Syntactic Variation: Particle Movement Revisited. *Journal of Quantitative Linguistics*, Vol. 8, No. 1: 33-50.

Ivanov, V. V. et al. (ed.) (1989). *Kategorija posessivnosti v slavjanskich i balkanskich jazykach*, Moskva.

Luce, R./Tukey, J. W. (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology*, 1: 1-27.