

CONSCIOUS WORD-FORMATION AND LINGUISTIC COMPETENCE

Moharram Eslami(1,2), Samaneh Abhari (1) Safura Fada'izadeh (1)

1. Zanzan University, 2. Technische Universität Dresden

Institut für Akustik und Sprachkommunikation

Moharram.Eslami@ias.et.tu-dresden.de

This paper is based on a corpus based research in which the texts with different subjects amounting to 500.000 words were transcribed phonemically by three people holding MA in linguistics during two years. This corpus includes only 30,000 words. Then the inflected forms and foreign words were deleted manually and the output was a lexicon including 16883 lexemes. These lexemes had different frequencies in our corpus and their total weight was equal to 419082 lexemes. The lexemes were divided into seven classes according to the number of syllables in their structure. The data were analyzed from the view point of absolute frequency and total frequency to reach to the most common and unmarked syllable structures in Persian in order to use the results in evaluating the conscious word-formation which is currently done by The Academy of Persian Language and Literature (APLL). Every class of the lexemes were analyzed according to the type, number and arrangement of their syllables to find out the most unmarked structures within every specific class, for example, in two syllable words, the syllable structures CV.CVC and CVC.CVC have the most frequencies respectively as in Persian words *madar/mother* and *jadval/table*. Therefore it is justified that one newly made two syllable word having the above structures may be accepted easily than the word having the structure of CVCC.CVCC like the Persian equivalent *barqkaft* for the English word *electrolysis* by APLL. If the high frequency could be considered as a symbol of unmarkedness in languages, then one can conclude that word-formation in the framework of unmarked structures is in accordance with the native speakers' linguistic competence. Relying only on the frequency, and disregarding other linguistic criteria such as syntactic and semantic limitations, one can say that on the basis of Persian phonotactics and phonological system, between two options like "*balgard*" and "*charxbal*" for the English word "*helicopter*", "*balgard*" has more chance to be accepted by the natives and the experience supports this claim. These two equivalents have the syllabic structures of CVC.CVCC and CVCC.CVC and these structures have the frequencies of 230 and 100 respectively in our data.

This research investigates the three books which include the equivalents presented by APLL for borrowed words and concepts and compares it with the results of the research on natural data to show the degree of accordance of their efforts with linguistic competence of the natives. Of course the writers of this paper are aware of the limitations and scientific considerations in specialized word-formation, but our data emphasize that those syllable structures which are absent or have little frequencies in natural data are hard to be accepted by natives.

This research suggests that in conscious word-formation as a kind of language planning, the linguistic intuition of the natives must be taken into account; otherwise any attempt in this relation is subject to failure.