DOCUMENTS ET COMMUNICATIONS

THE HYPOTHETICAL STANDARD-LENGTH OF WORDS IN ARABIC AND ITS REALIZATION IN THE VOCABULARY

G. Herdan in his LCC [1], TTM [2], and CLO [3] discusses the problem of the word-lengths in languages and their dependance on the quantity of words and phonems (letters) in the given language. In order to calculate the standard word-length in a given language he applies the following equation from the conbinatorial mathematics:

$$N = n(n^r-1)/n-1$$

in which N corresponds to the quantity of words in the given language, n to the quantity of phonemes (letters) and r represents the word-length.

The aim of the present paper is the verification in what extent the above-mentioned probability law has been realized in the Vocabulary of the Arabic language. For this purpose the above-inserted equation has been applied. As the basis of word-number (N) serve 100.000 words included in the Arabic Dictionary: *Lisān al-ʿarab* [4]. As phonems (n) to which the letters of the Arabic alphabet, because of

---

[1] Gustav Herdan, *Language as Choice and Chance*, Groningen 1956, pp. 131—132; later on referred to as LCC.

[2] Gustav Herdan, *Type-Token Mathematics*, S.-Gravenhage 1960, p. 37.

[3] Gustav Herdan, *The Calculus of Linguistic Observation*, S.-Gravenhage 1962, p. 20.

[4] Ǧamāl ad-dīn al-Anṣārī, (1232—1311), *Lisān al-ʿArab*, s. GAL, II, pp. 21—22, S. II, pp. 15—16.

their phonetic character, are exceptionally adequate, have been regarded:

Consonants						Vowels	
' ء						Short	Long
b ب	z ز	f ف				a ـَ	ā ـَا
t ت	s س	ḳ ق				i ـِ	ī ـِي
ṯ ث	š ش	k ك				u ـُ	ū ـُو
ǧ ج	ṣ ص	l ل				ـً un, ـٍ in, ـً an, = auslaut	
ḥ ح	ḍ ض	m م					
ḫ خ	ṭ ط	n ن					
d د	ẓ ظ	h ه					
ḏ ذ	ʿ ع	w و					
r ر	ġ غ	y ى					

They amount to 35 altogether. After having substituted the N and n with these figures, the above-mentioned equation presents itself in the following way:

$$100.000 = 35(35 - 1)/35 - 1$$

$$r \simeq 3 - 4$$

The resulting figures have been confronted with the words in two dictionaries, viz. that of Wehr [5] and that of Baranow [6], what is shown in the below-inserted table followed by the diagram (see p. 311). The sampling of words has been perfomed according to the random sampling principles exposed in Herdan's LCC [7].

*

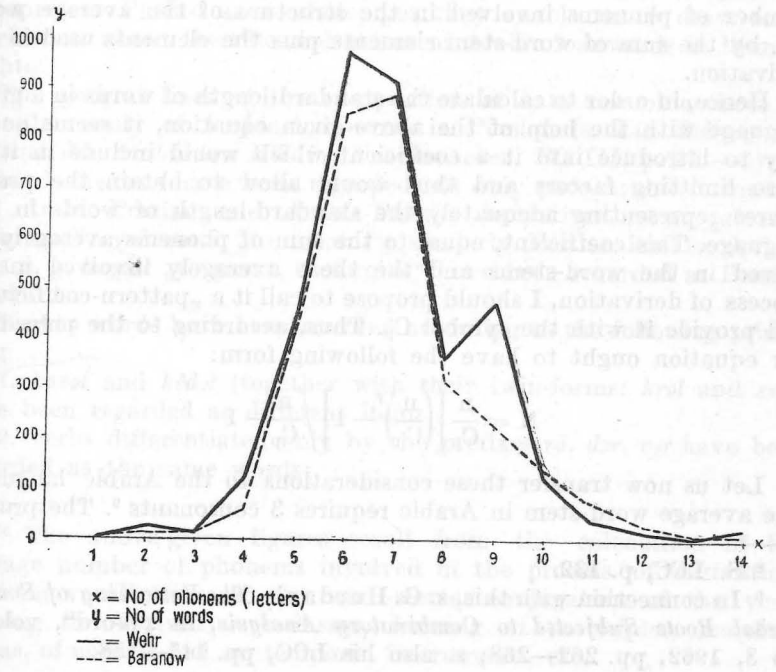
As it appears from the above-given data, the length of Arabic words oscilates between 2 and 14 letters. The lack of 1-letter long words is easily explicable, since in Arabic no syllable may have a vowel in its anlaut. In both dictionaries, most copiously represented are 6- and 7-letters long words, hence, they are to be regarded as representing the standard-length of Arabic words. This in comparison with the above-obtained $r \simeq 3 - 4$, leads to the conclusion that the effective demand on phonems in the greater part of Arabic words and hence their legth show a great redundancy. This redundancy however is apparent. One ought namely to keep in mind that the above equation would be fully applicable to this kind of calculations if, in a given

[5] Hans Wehr, *Arabisches Wörterbuch für die Schriftsprache der Gegenwart*, 2 Bd., Leipzig 1952.

[6] Ch. K. Baranow, *Arabsko-russkiy słowar'*, Moskwa 1957.

[7] S. pp. 26—30.

No of phonems (letters)	Wehr		Baranow	
	No of words	%	No of words	%
1	—	—	—	—
2	16	0·4	1	0·03
3	1	0·03	4	2·1
4	121	3·4	64	9·0
5	449	12·8	432	13·3
6	968	27·6	847	25·9
7	911	26·0	884	27·1
8	357	10·2	317	9·7
9	476	13·5	469	14·4
10	140	4·0	143	4·4
11	45	1·3	67	2·0
12	14	0·4	27	0·8
13	—	—	4	0·1
14	6	0·2	4	0·1
	3504		3266	



language, all combinations of phonems were admissible, and which, as it is commonly known, are not [8], since words of a given language follow in their structure the pattern depending on the distribution of phonems valid in this language and the total liberty of the combinations of phonems may only be assumed at its „birth", since words which were coming into being in later stages had, if they were to fall within the lexical system, to adapt themselves to the meanwhile developped laws hich became obligatory for this language. Thus, the length of words in the language is conditioned not only by the number of words, the number of phonems (letters) as well as by laws of their distribution but also by the average quantity of phonems involved in the structure of one particular average word. All these just mentioned factors must influence reducingly the effective quantity of phonems staying at the disposal of their possible combinations leading to the creation of words which would fall within the lexical system of the given language, since the word-stems in this language involve such and such average number of phonems and the derivatives from them involve again certain average number of them, all arranged in a more or less strictly determined way.

After having considered all this what has been said, it can easily be seen, as it seems, that at the n-number of phonems, their possible quantity of combinations will be diminished by the whole average number of phonems involved in the structure of the average word, i. e. by the sum of word-stem elements plus the elements used in the derivation.

Hence, in order to calculate the standard-length of words in a given language with the help of the above-given equation, it seems necessary to introduce into it a coefficient which would include in it all these limitting factors and thus would allow to obtain the proper figures representing adequately the standard-length of words in this language. This coefficient, equal to the sum of phonems averagely involved in the word-stems and the those averagely involved in the process of derivation, I should propose to call it a „pattern-coefficient" and provide it with the symbol C_p. Thus, according to the preceding, our equation ought to have the following form:

$$N = \frac{n}{C_p}\left[\left(\frac{n}{C_p}\right)^r - 1\right] \Big/ \frac{n}{C_p} - 1$$

Let us now transfer these considerations to the Arabic language. The average word-stem in Arabic requires 3 consonants [9]. The process

[8] S. LCC, p. 132.

[9] In connection with this, s. G. Herdan's, *The Patterning of Semitic Verbal Roots Subjected to Combinatory Analysis*, in „Word", vol. 18, No 3, 1962, pp. 262—268; s. also his LCC, pp. 145—158.

of derivation involves averagely 3 additional phonems [10]. Thus, the coefficient C_p for the Arabic will be equal to the sum: $3 + 3 = 6$. We come now to the calculation of r with the help of the modified equation:

$$100.000 = \frac{35}{6}\left[\left(\frac{35}{6}\right)^r - 1\right] \Big/ \frac{35}{6} - 1$$

$$r \simeq 6, -7$$

The obtained result representing the hypothetical standard-length of words in Arabic fully corresponds with the actual state confirmed by the Dictionaries as it has been proved by the above-inserted table and diagram.

A. Czapkiewicz

A NOTE ON PERSIAN AND ARABIC LOANWORDS IN PUSHTO

Presented statistical data have been obtained from a relatively small sample of Khushhal Khan Khatak's poems and cannot pretend to represent Pushto as a whole; yet they enable us to draw some conclusions about the rôle and status of Arabic loanwords in literary Pushto.

The general lines of the compilation of data resemble those in a similar study of Arabic loanwords in Persian (s. *Ein Beitrag zur Statistik der Arabischen LW im Neupersischen*, FO III, p. 171—5): as Arabic loanwords have been regarded not only pure Arabic forms but also Arabic-Persian and Arabic-Pushto derivatives, e. g. *χabardār* „informed", *χabaravəl* „to inform"; similarly Persian-Pushto derivatives e. g. *baχšəl* „to grant, to forgive" have been regarded as Persian loanwords.

Pushto words have been counted according to the following principles:

1. *kavəl* and *kēdəl* (together with their twin-forms: *krəl* and *svəl*) have been regarded as different items;

2. verbs differentiated only by the prefixes *rā*, *dər*, *vər* have been regarded as the same words;

[10] The above-given figures result from the calculation of the average number of phonems involved in the process of derivation, viz.: the vocalization of the word-stems, conjugation of the verb, forming the derived verb-classes, forming of participles, maṣdars, nisbas, of nomina loci ac temporis, instrumenti, etc.