

**DESCRIPTOR POTENTIAL OF THE UKRAINIAN LANGUAGE FROM
THE PERSPECTIVE OF NEW METHODS OF LEXICAL ANALYSIS OF
PERSONALITY STRUCTURE**

<https://doi.org/10.37096/SHDISJ-20-2.2-0004>

Gorbaniuk Oleg

ORCID ID: 0000-0001-9830-8537

Bevz Galyna

ORCID ID: 0000-0003-2487-5429

Gorbaniuk Julia

ORCID ID: 0000-0001-7732-7819

Krasa Pavel

ORCID ID: 0000-0001-8153-0318

*Possibilities of analyzing Ukrainian personality lexicon in the light of new
method of psycholexical research*

Abstract

Introduction. The relevance of the study is caused by the existing need to develop psycholexical research, which will allow the universalization and operationalization of theoretical psychological constructs and models. The specifics of psycholexical research are revealed. The main advantages and limitations of the psycholexical method are analyzed.

Methodology. Prospects for the analysis of the Ukrainian language with the help of a new method of separating the factor structure of low and high dimensions are presented. The initial provisions and procedure of psycholexical research on the basis of the Ukrainian lexicon of personality are described.

Results. Based on the determination of the appropriate number of factors and the comparison of different methods of data processing, the choice of personality descriptors and the rotation of the factors, structures of different dimensions are singled out on the basis of the same data.

Discussion. The authors emphasized on the need to analyze the resulting structures, taking into account the theory of the psycholexic industry, as well as giving theoretical meaning to open dimensions, checking their descriptive and explanatory meaning on the way to creating a theoretical model.

Keywords: personality structure, psycholexical method, Ukrainian language, descriptive potential of language.

Introduction

Achieving a taxonomic compromise is important for any science (for example, Darwin's or Mendeleev's taxonomy). Acceptance of the taxonomy of the subject of research contributes to the development of the industry due to the unambiguous terminology and to

acceptance of the same models and universalization of operationalization of theoretical terms. The task of psychology – as a science that aims to explain human behavior - includes a systematics of individual differences, and the key are differences in personality traits. The research method that allowed achieving

taxonomic agreement in the psychology of personality is a psycholexical method. Its greatest achievement is the model of the Big Five Personality Traits, which is widely accepted in the modern scientific and psychological environment (Saucier & Srivastava, 2015).

The starting point for psycholexical research is the lexical assumption, according to which:

1) All the most important differences for the social functioning of man were coded in natural languages;

2) The more terms a given nation/culture have created for a given difference/trait, the more important a given personality trait is;

3) We can identify the most important features while analyzing the structure of the lexicon of personality (Galton, 1884; Goldberg, 1981).

An example of a lexical assumption about the encoding of personality traits in natural languages is presented below in Figure 1.

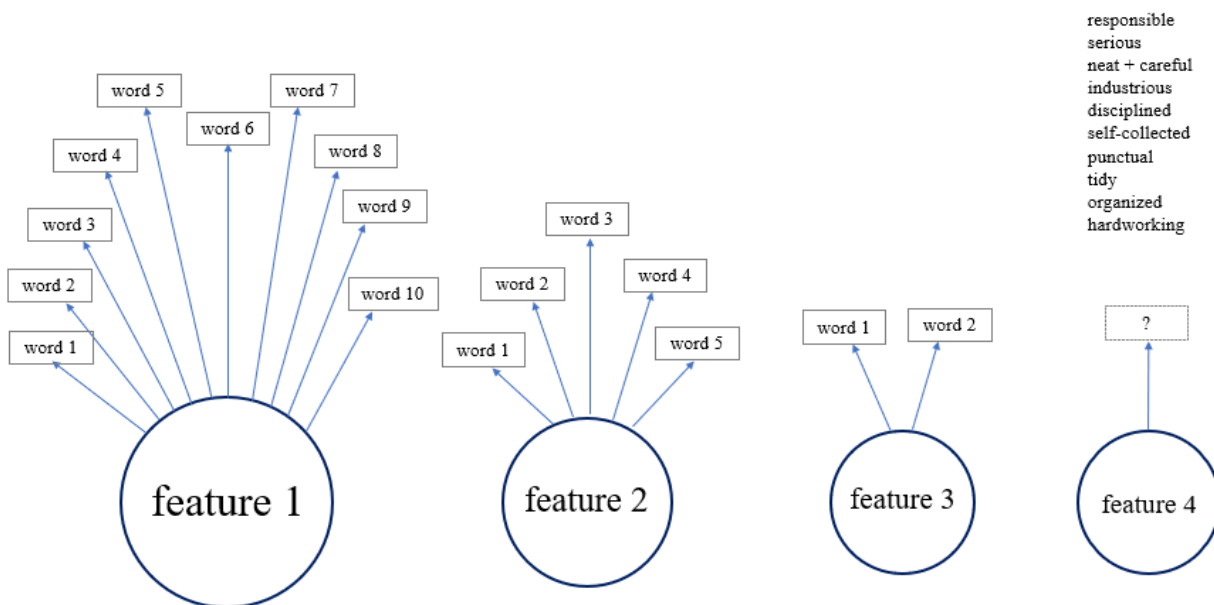


Fig.1 Lexical assumption about the encoding of personality traits in natural languages.

Modern psycholexical research over the past 30 years has covered about 30 languages out of more than 5-6 thousand languages in the world (De Raad & Mlacic, 2020). Among the many conclusions that can be drawn from the results of these studies, we would like to highlight a few key ones: 1) the lexicon of personality has a hierarchical structure of the organization, 2) in the case of higher levels of abstraction, personality traits are more universal for many languages, 3) 5-6-factor models explain no more than 18-25% of individual differences that can be

described in languages, 4) the methodological shortcoming of research is the lack of criteria that would indicate the optimal structure (factorial decision), which is the reason for differences in research results, regardless of whether such differences actually exist between languages / cultures / nations.

The aim of the research. Ukrainian is one of the languages for which the structure of the personality lexicon has not yet been established (Gorbaniuk et al., 2018). The aim of the conducted psycholexical research was to establish

the structure of the Ukrainian lexicon of personality with low and high dimensions.

Methodology

Sample. The sample of the empirical study consisted of 835 Ukrainian-speaking students from universities in Ukraine and Poland. They made a self-description of their personality on the basis of the Ukrainian list of human qualities created by the authors of the research. The task of the respondent in the process of research was to assess with the help of 7-point scale, how exact each word from the list accurately characterizes him as a person. If the respondent did not understand the

meaning of the word, it was necessary to put "0".

All lexical research can be reduced to two most important stages (Gorbaniuk & Ivanova, 2018): 1) qualitative analysis of the vocabulary of natural language, the purpose of which is to identify the complete vocabulary of the individual, 2) quantitative research aimed at identifying the factor structure of the personality lexicon based on the use of the lexicon by its users in the process of self-description and description of the characteristics of other people. In graphical form, the main stages of psycholexical research are presented in Figure 2.

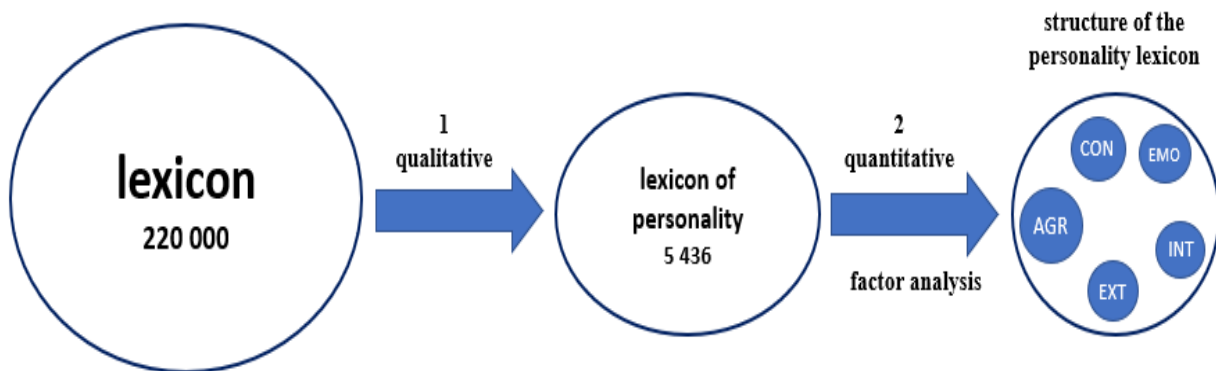


Fig. 2. Stages of psycholexical research

Method. The authors made the selection and classification of words describing human characteristics using German taxonomy (Angleitner et al., 1990). The word selection process based on the Ukrainian dictionary (220,000 words) is described in details in a separate article (Gorbaniuk et al., 2018). The final list of terms describing personality (including terms that are highly valued) contains 1053 adjectives. Selected words are the best representatives of morphemes with very similar meanings and they are understandable to more than 90% of students. The exception made protocols in which 1) there was a poor knowledge of words by the respondent (>10%), 2) the absence in the form of a list of specific respondents filled in data (>10%). The method Input Forest was used in other forms to fill in the missing answers.

Saucier & Iurino (2020) developed a proposal to identify structures with an increased number of measurements, that complements existing structures, which consist of one to six to seven factors. The procedure proposed by the authors allows minimizing the arbitrariness of the choice of factor structure with the highest degree of detail for a given language, resulting from the various structures obtained on the basis of: 1) original vs ipsatized data (raw vs ipsatized data), 2) orthogonal vs oblique

rotations (Equamax vs Oblimin rotation), 3) data of self-description in comparison with the description of other people's personality (self-vs peer-rating). This gives a total of 8 combinations of possible results: data x rotation x perspective. Exploratory factor analysis (EFA) is performed separately for each of the 8 combinations.

At the first stage of factor analysis, Parallel Analysis is performed in order to identify the maximum possible number of non-random k factors for each data type.

In the second stage, the EFA is calculated cyclically until three conditions will be met simultaneously by systematically reducing the number of factors, starting with k factors (See Table 1):

| r | • with each of the factors correlates at least 8 personality descriptors with a load factor > 0.30 ;

| r | • at least one adjective correlates with the factor > 0.40 .

• the factor has a clear and consistent psychological interpretation.

Table 1. Number of Factors Recommended by Parallel Analysis and the MAP Procedure and Candidate Models Derived from Each Dataset Variant Having All Factors of Sufficient Size and Interpretability

data	object	Number of Factors Recommended By		Solutions Having All Factors of Sufficient Size (at least one item $r > 0,40$ + at least 8 items $r > 0,30$) and Interpretability	
		PA	MAP	Orthogonal (Eqx)	Oblique (Obl)
ips	self	41	>41	19	21
raw	self	34	>34	20	28

* numbers of factors reduced until a solution with a full set of sufficiently sized and interpretable factors was found. Ips. – ipsatized data. Orig. – original data. Eqx – equamax. Obl – oblimin ($\delta = 0$).

Perspective for each number of factors, which corresponds to all criteria from the second stage of the research taken together. This similarity between factor structures is estimated using Tucker coefficients (for different perspectives), Pearson correlation (for different rotations) and canonical analysis (original vs ipsatized data). It is necessary to choose a factor solution \times rotation \times In the third stage, the coherence coefficients between the structures are calculated for all eight combinations (data n and

perspective) reach the highest value. We can also define several competing solutions, the advantage of one of which will be established only in the following stages. \times rotation \times for it the similarity coefficients between the structures for all eight combinations (data

In the fourth stage, the best 8 markers/descriptors for each of the n factors for each combination are identified (data x rotation x perspective). Optimally 4 markers/descriptors represent each extremity of the continuum. Then

Cronbach's alpha reliability coefficients are calculated for each of the n scales. In such way the psychometric potential of the future operationalization of the personality model is established.

In the fifth stage, the explanatory value of the established structures in relation to the behavior is calculated. In order to do this, multiple correlation coefficients R between the n -factor structure and human behavior indicators are calculated.

In the sixth stage, the optimal solution is selected for each of the competing n -factor solutions. It is based on a comparison of the reliability of the scales and multiple correlation coefficients. This means that the final structure is: 1) the most resistant to differences arising from the use of different types of data, rotation and perspective of the description of personality, 2) is the most detailed in describing personality traits in individual dimensions, 3) is with the greatest psychometric potential in the process of operationalization, 4) is with the greatest ability to predict human behavior.

Results

The authors distinguish linguistic structures of personality of low and high dimension on the basis of the same data, by using the method of determining the appropriate number of factors and comparing different methods of data processing (original vs ipsatized), different methods of choosing personality descriptors (with and without highly evaluative terms) and rotation of factors (orthogonal vs oblique).

In the case of a low-dimensional structure, the two-factor solution corresponded to the "big two", but the three-, five-, and six-factor structures did not at least partially overlap with the universal models of the "big three," "big five," or "big six, when the list contained highly valued adjectives. Methods for determining the maximum number of factors recommended more than five or six factors. We compared 8 candidate structures obtained from different combinations of methods. We compared 8 candidate structures obtained from different combinations of methods. Also we defined the 21-factor structure as the most reliable and promising model compared to all other models (See Table 2).

Table 2. Robustness Indices for the 4 Candidate Models

Candidate Models (Origin)	N of factors	Average Orthogonal-Oblique Best – Match Correlation		Canonical analysis	
		lps	Orig	lps vs Orig (Eqx)	lps vs Orig (Obl)
19 factors, lps, Self, Eqx	19	0.831	0.760	0.817	0.847
20 factors, Orig, Self, Eqx	20	0.770	0.797	0.827	0.853
21 factors, lps, Self, Obl	21	0.840	0.800	0.823	0.851
28 factors, Orig, Self, Obl	28	0.819	0.784	0.820	0.847

The comparison was made on the basis of maximum coefficients of agreement, thoroughness, factor independence and high coefficients of predicted validity of human behavior (See Table 3).

Table 3. 21-Factor Structure (21 factors, Obl, Ips), personality-adjectives with highest factor loadings

N	Factor	Альфа	Personality-adjectives
1	industriousness	0.762	capable of working successfully; feasible, thrifty, responsible, dutiful, industrious, reliable, hardworking, operative, efficient, practical, housewifely
2	Talkativeness	0.886	taciturn, silent, tacit, reticent, unconversable, taciturn, quiet, chatty, talkative, conversational
3	apprehensiveness	0.792	apprehensive, weepy, maudlin, cowardly, anxious, apprehensive, indecisive, hesitating, brave, self-confident, perky, assured
4	lack of intelligence	0.816	stupefied, empty-headed, moron, imbecile, stupid, foolish, oakhead, backward, retarded
5	Organization	0.740	organized, tidy, neat + careful, punctual, disciplined, lazy, disorderly, serious
6	Helplessness	0.711	uneducated, unreasonable, unconscious, ignorant, not well-read, infirm, absent-minded, helpless, slow-witted
7	Stubbornness	0.591	who likes to slander, but also loves to speak wittily; stubborn, controversial, unyielding, principled, naughty, weak character, non-persistent
8	Craziness	0.835	insane, nuts, crazy, doltish + crazy, gaga, dejected, depressed
9	Creativeness	0.703	literate, creative, poetic, savant, gifted, talented, art critic
10	Mercenary	0.694	self-seeking, calculating, interested, self-serving, not mercenary, vulgar, slippery, malicious, adhesive, selfless, self-denying
11	submissiveness, humility	0.662	complaisant, obliging, compliant, suggestible, someone, who like to help others, sacrificial, humble, submissive, obedient, submissive, obedient, susceptible-compliant
12	Heroism	0.806	chivalrous, heroic, courageous, fearless, strong, tough, brave, intrepid
13	imbalance, unpredictableness	0.717	mad, wild, unbridled, mischievous, misbehave, riotous + exuberant, frantic, unbalanced, audacious, capricious

14	Optimism	0.806	enjoying life, optimistic, who loves life (optimistic, cheerful, vivacious), joyful, gloomy, depressive, pessimistic, nervous
15	Flexibility	0.587	unstable, impervious, unresponsive, not flexible, not deep, not quick, aplastic, inelastic, quarrelsome, snappish, not appeasable, non-sagacious
16	Dynamism	0.778	fast, energetic, nimble, clever, quick, inactive, not fast, unhurried
17	Honesty	0.709	deceitful, hypocritical, two-faced, crafty, mendacious, corrupt, truthful, veracious
18	quarrelsomeness	0.741	rude, fierce, sharp-tongued, hard, irritable, scandalous, swearing, gruff
19	self-determination	0.650	far-sighted, self-defined, self-identified, practical, sighted, conscious, enthusiast, cautious, ambitious
20	haughtiness, egoism	0.757	selfish, proud, arrogant, power-hungry, narcissistic, dictatorial, authoritarian, self-praise, domineering
21	Sensitiveness	0.790	mild-tempered / meek, gentle, delicate, romantic, sensitive, affectionate, soft, pleasant, kind-hearted, softhearted.

Below are the results of multiple regression analysis (See Table 4)

Table 4. Multiple Correlations for Predicting Self-Reported Variables from Factors of Candidate Models

	Predicted / explained variable	Data:	Ipsatized								Original			
			x	x	x	x	x	x	qx	bl	x	x	qx	bl
		Rotation:												
		Factors:							9	1			0	8
	Happiness	R	.39	.49	.49	.60	.60	.63	.77	.77	.56	.56	.73	.79
	Number of acquaintances	R	.22	.42	.44	.44	.45	.45	.57	.57	.41	.41	.50	.59
	Number of (close) friends	R	.03	.04	.06	.14	.16	.17	.32	.37	.20	.20	.36	.42
	Constant sexual partner	R	.02	.18	.22	.25	.31	.32	.62	.66	.43	.42	.66	.69
.1	Playing computer games	R	.01	.03	.17	.18	.33	.33	.59	.63	.20	.39	.66	.77
.2	Communicating	R	.13	.21	.28	.29	.30	.35	.44	.53	.26	.28	.47	.60

	on social media													
.3	Browsing websites or movies	R	.31	.40	.42	.42	.45	.45	.61	.63	.35	.37	.57	.66
.1	Consuming strong alcohol >20%	R	.11	.18	.18	.19	.17	.17	.36	.37	.20	.20	.35	.37
.2	Consuming medium alcohol 9-20%	R	.04	.10	.30	.30	.30	.35	.42	.44	.35	.35	.48	.58
.3	Consuming weak alcohol 1-8%	R	.05	.05	.18	.19	.26	.28	.45	.46	.22	.35	.51	.58
	Smoking	R	.09	.09	.11	.11	.10	.24	.50	.52	.29	.29	.60	.65
	President elections	R	.38	.38	.39	.43	.46	.46	.62	.63	.44	.47	.60	.73
	Parliament election	R	.36	.37	.37	.37	.39	.45	.65	.66	.42	.44	.58	.73
0	Membership of organizations (social/political activity)	R	.15	.28	.28	.32	.33	.36	.47	.48	.32	.32	.46	.55
1	Sports exercises	R	.06	.33	.44	.52	.53	.53	.63	.66	.50	.51	.62	.68
2	School scores	R	.23	.24	.24	.48	.48	.52	.73	.75	.50	.51	.73	.77
3	BMI	R	.07	.09	.17	.18	.22	.22	.52	.55	.37	.37	.53	.63

Note. N=90. R – multiple correlation. Adj. R – adjusted R value, with any negative values replaced by .00. Vx. – from varimax rotation. Eqx. – from equamax rotation. Obl. – from oblimin rotation. Q – Questions.

The figure shows the hierarchical structure of the lexicon of the low-dimensional personality.

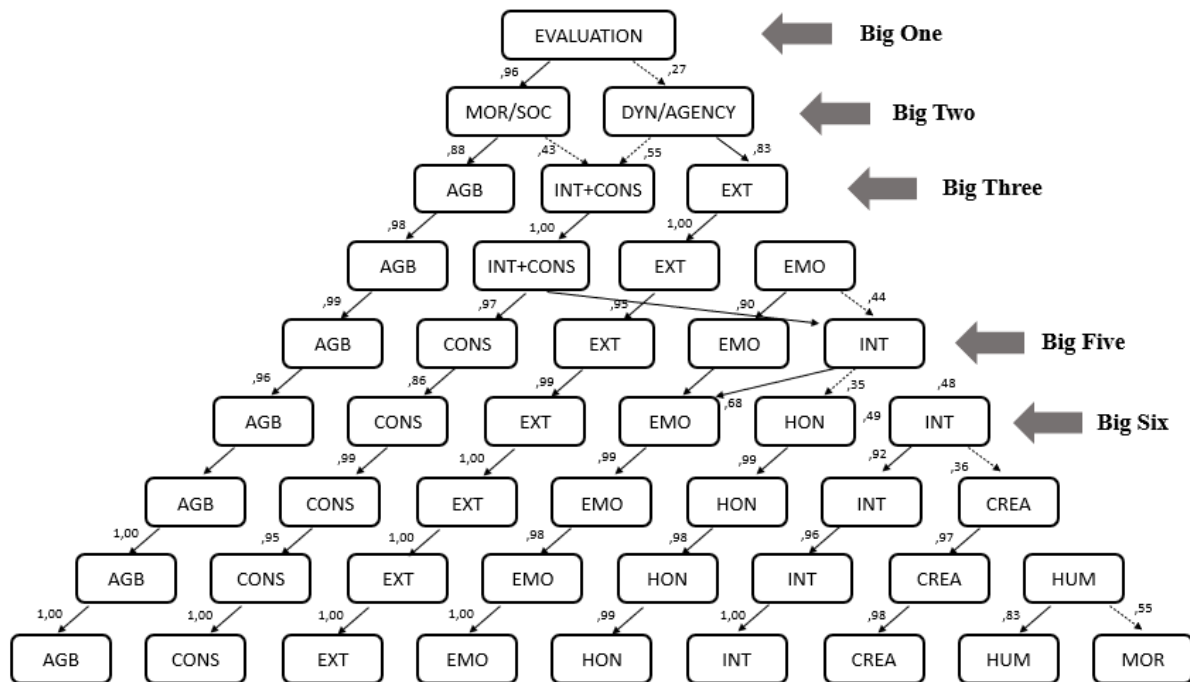


Fig.3. Hierarchical structure of the lexicon of low-dimensional personality.

Conclusions

The lexical approach is a valuable alternative in modern scientific and psychological studies of human perception due to its unique features. The experience of lexical analysis in the process of development and systematization of individual differences gives a chance to reach an agreement between researchers on the scientific systematics of the studied phenomena and their methodological research with the help of compatible models and research tools developed on the basis of the results of lexical research.

In addition to possibilities, the lexical approach also hides dangers when the result of lexical research in the form of lexicon structure is perceived literally as a theoretical model. It is assumed that the naive observer will notice and name all aspects of the phenomenon, in which the

researcher is interested in. Before becoming a model, the chosen structure should be analyzed, taking into account the theory of this field of knowledge. And only after giving theoretical meaning to open dimensions and checking their descriptive and explanatory meaning, it can become a theoretical model in the full sense of the term. Therefore the result of lexical research is only an inspiration for the development of the theory, but it cannot be its replacement.

References

- Angleitner, A., Ostendorf, F., & John, O. P. (1990). Towards a taxonomy of personality descriptors in German. A psycho-lexical study. *European Journal of Personality*, 4, 89–118.
- De Raad, B., & Mlačić, B. (2020). The Big Five Personality Trait Factors. In

- Oxford *Research Encyclopedia of Education*. Oxford Press.
- Galton, F. (1884). Measurement of character. *Fortnightly Review*, 36, 179–185.
- Goldberg, L. R. (1981). Language and individual differences. The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141–165). Beverly Hills, CA: Sage.
- Gorbaniuk, O., Ivanova, A. (2018). Język jako źródło wiedzy o różnicach indywidualnych. *Nauka*, 4, 7-26
- Gorbaniuk, O., Mirowich, A., Leoszko, W., Gorbaniuk, J., Kordon, A., Świdorska, M., Kuts, O., & Korczak, A. (2018). A psycholexical classification of Ukrainian descriptors of individual differences. *Current Problems of Psychiatry*, 19, 1–8.
- Saucier, G., & Iurino, K. (2019). High-dimensionality personality structure in the natural language: Further analyses of classic sets of English-language trait-adjectives. *Journal of Personality and Social Psychology*, 119(5), 1188–1219.
- Saucier, G., & Srivastava, S. (2015). What makes a good structural model of personality? Evaluating the Big Five and alternatives. In M. Mikulincer & P. R. Shaver (Eds.), *APA Handbook of Personality and Social Psychology*, Vol. 4: Personality Processes and Individual Differences (pp. 283–305). New York: Cambridge University Press.