UNIVERSIDAD & SOCIEDAD
Presentation date: February, 2025
Date of acceptance: June, 2025
Publication date: June, 2025

PSYCHOLINGUISTIC

METHODS IN DATA PROCESSING

MÉTODOS PSICOLINGÜÍSTICOS EN EL PROCESAMIENTO DE DATOS

Zuleykha Baghirzadeh Murad [1*]
E-mail: zuleykha.baghirzada@au.edu.az
ORCID: https://orcid.org/0000-0002-4743-2893
Zulfıyya İsmayıl [2]
Email: ismayilzulfiyye@yahoo.com
ORCID: https://orcid.org/0000-0002-4967-0123
Mammadova Zeynab Ramiz [2]
Email: mammadovazeyneb@ndu.edu.az
ORCID: https://orcid.org/0009-0008-5737-9935
Jeyhun Aliyev Ramiz [2]
Email: ceyhuneliyev@ndu.edu.az
ORCID: https://orcid.org/0000-0002-4291-3989
Ruhangiz Aliyeva Mammad [3]
Email: ruhangiz.aliyeva@list.ru
ORCID: https://orcid.org/0000-0002-9466-4822
Nigar Aliyarova Namig [4]
Email: nigar.aliyarova@baau.edu.az
ORCID: https://orcid.org/0009-0006-5580-7903
[1]Azerbaijan University, Azerbaijan.
[2]Nakhchivan State University, Azerbaijan.
[3]Nakhchivan Institute of Teachers. Azerbaijan.
[4]Baku Eurasian University, Azerbaijan.
*Author for correspondence

Suggested citation (APA, seventh ed.)

Baghirzadeh, Z. M., İsmayıl, Z., Zeynab, M. R., Aliyev, J. R., Aliyeva, R. M., & Aliyarova, N. N. (2025). Psycholinguistic Methods in Data Processing. ***Universidad y Sociedad, 17***(3). e5253.

ABSTRACT

Psycholinguistics, as an interdisciplinary discipline that bridges psychology and linguistics, plays a crucial role in understanding processes related to language and cognition. However, despite its relevance, there is a significant gap in the application of accurate and robust experimental methods that provide reliable results. This study aims to offer a review of the most widely used experimental methods in modern psycholinguistics, highlighting their advantages and limitations in the analysis of linguistic material. The findings suggest that the implementation of mathematical and statistical data processing techniques is essential for validating indicators derived from psycholinguistic experiments. These implications suggest that by adopting a more rigorous and systematic approach to methodology, researchers can improve the reliability of their results and contribute to more significant progress in the field of psycholinguistics.

Keywords: Psycholinguistics, Experiment, Reliability of results, Statistical and mathematical methods of data processing.

RESUMEN

La psicolingüística, como disciplina interdisciplinaria que une la psicología y la lingüística, juega un papel crucial en la comprensión de los procesos relacionados con el lenguaje y la cognición. Sin embargo, a pesar de su relevancia, existe una brecha significativa en la aplicación de métodos experimentales precisos y robustos que proporcionen resultados

fiables. Este estudio tiene como objetivo ofrecer una revisión de los métodos experimentales más utilizados en la psicolingüística moderna, destacando sus ventajas y limitaciones en el análisis del material lingüístico. Los hallazgos sugieren que la implementación de técnicas de procesamiento de datos matemáticos y estadísticos es esencial para validar los indicadores derivados de los experimentos psicolingüísticos. Estas implicaciones sugieren que, al adoptar un enfoque más riguroso y sistemático en la metodología, los investigadores pueden mejorar la fiabilidad de sus resultados y contribuir a un avance más significativo en el campo de la psicolingüística.

Palabras clave: Psicolingüística, Experimento, Fiabilidad de resultados, Métodos estadísticos y matemáticos de procesamiento de datos.

INTRODUCTION

When reading any scientific work, be it a monograph, an article or a thesis, two questions come to the fore: what is the research material and what methods are used in its analysis? If the material and methodology do not correspond to the set goals and the topic under consideration, the results and conclusions obtained cannot be considered reliable. In this regard, psycholinguistics is the science of speech production and perception, and within this discipline language is studied in relation to speakers and their activities. In the local tradition

A. N. Leontiev, E. F. Tarasov, Yu. A. Sorokin, etc., psycholinguistics is interpreted as a theory of speech activity. "The subject of psycholinguistics is the relationship of the personality with the structure and functions of speech activity, on the one hand, and language as the main 'shaper' of the image of the human world, on the other" (Leontiev, 2003). Scientists note that the focus of attention of the science of psycholinguistics is the thought processes involved in language acquisition and use, which are well established and are completely reviewed from the point of view of the practical application of linguistic phenomena at all levels (phonetics, vocabulary, grammar) (Slobin, 1972).

Language mediates the functioning of all human cognitive mechanisms: perception of the surrounding reality, processing of information, formation of conscious images and awareness of the individual as a member of a certain linguistic and cultural community. This understanding of psycholinguistics is based on various aspects of anthropocentric knowledge of language and has shaped the modern appearance of this discipline. Among many of them, we will first mention the assumption about the connection between language and mental activity. This assumption was expressed by von Humboldt (2000) in the form of the idea of the existence of the internal form of the language, by Potebnya (1999) in the form of the idea of the internal form of the word. It is difficult to imagine the development of psycholinguistics without the theory of linguistic relativity of Sapir (1949) and Whorf (1956), the psycho-sociological theory of Mead (1962) and the theory of historical ethnology of Boas (2021). The central place in the cultural-historical theory of Vygotsky (2001) is the idea of the inseparability of linguistic processes and thought.

Along with the formulation of theoretical provisions of a new scientific direction, the subject and object of study are always defined, research methods are developed, and the boundaries of the discipline are outlined. Psycholinguistics exists in an interdisciplinary field, and therefore it is often difficult to "draw the line where other sciences begin" (Klyukanov, 2018). Psycholinguistics as a scientific discipline is unique in that, despite the existence of clearly defined scientific boundaries, it has its own established experimental fund, but represents an open set. As L.S. Vygotsky rightly stated, the problem of thinking and speech is one of the most complex, the study of which includes "the study of experimentally established concepts, the study of written speech and its relationship to thinking, the study of inner speech, etc." This can be achieved through a number of specific experimental studies (Vygotsky, 2001).

In this regard, psycholinguistic methods include experimental techniques such as reaction time measurement, priming paradigms (pre-activation), and speech error analysis, which are designed to understand processes like language comprehension, production, and acquisition. These techniques, originally developed to study human cognition, have been adapted to data processing to analyze large volumes of text or discourse elements, particularly in the current context of natural language processing (NLP) and even in the development of artificial intelligence systems such as large language models (LLMs) capable of generating and understanding human language. Data processing, in this case, involves managing and analyzing linguistic data, such as text corpora or speech transcriptions. However, given recent advances, these techniques can extract insights into how language models represent and process language and how they align with human processes.

Recent research suggests that psycholinguistic methods are crucial for improving LLMs by testing their ability to reflect human cognitive processes. For example, a recent study by Duan et al. (2024) employs psycholinguistic paradigms, such as sound-gender association tasks and implicit causality, to explore neuron-level representations in the GPT-2-XL model. This study reveals that while the

model struggles with tasks like sound-shape association, it demonstrates human-like abilities in others, such as implicit causality, suggesting that certain neurons specialize in specific linguistic competencies. Another example is the work of Futrell et al. (2019), which uses psycholinguistic methods to test recurrent neural network (RNN) models, such as LSTMs, in their ability to represent incremental syntactic states. The authors found that LSTMs trained on large datasets represent syntactic states comparably to more specialized models like RNNGs, helping refine how these models process complex linguistic data, such as sentences with intricate syntactic structures. These applications not only improve model accuracy but also provide a window into how LLMs mimic human cognition, which is essential for applications like text generation, machine translation, and dialogue systems.

A less conventional but equally significant application is the use of psycholinguistic tools in software engineering text analysis. Sajadi et al. (2025) conducted a systematic review on the use of LIWC (Linguistic Inquiry and Word Count), a psycholinguistic tool that analyzes texts to identify dimensions such as emotional, cognitive, and social content ("Psycholinguistic Analyses in Software Engineering Text: A Systematic Literature Review"). They found that LIWC is used to analyze communications like code reviews and chats, offering insights into developers' psychological states, such as stress or collaboration, which can enhance decision-making and team productivity. This application demonstrates how psycholinguistic methods extend beyond NLP, integrating into data processing in contexts where language reflects human states, expanding their utility in interdisciplinary research.

Nevertheless, while the use of psycholinguistic methods in data processing offers benefits such as greater interpretability of language models, aligning them with human processes, and the ability to uncover cognitive and emotional patterns in large datasets, it also faces challenges, such as the need for large amounts of labeled data and the difficulty of generalizing human experiment results to computational models, especially when models do not always replicate human behavior. With these considerations, this article briefly analyzes some important elements to use the most popular methods of modern psycholinguistic research, and it is presented possible methods for the mathematical and statistical processing of data obtained through psycholinguistic methodologies.

DEVELOPMENT

Determining the role of language in thinking processes, such as speech formation, speech production and speech generation, requires the participation of subjects. When forming a fund of experimental studies in psycholinguistics, the achievements of related sciences - linguistics and psychology - are taken into account. In this context, one of the important issues is the effectiveness of using linguistic and psychological methods in psycholinguistic research. Language material is a guide in the field of work of cognitive mechanisms, and language as a means of speech activity is used by psychologists to determine the norm or pathology in the psychological state of the subject. For example, psychological questionnaires and psychological experimental tests are used to analyze various aspects of personality psychology. The psychological technique of sentence completion according to the Sacks & Levy (1950) sentence completion model (SSCT) (1950) is suitable for studying the development of inner speech. We believe that psycholinguistic experiments aimed at assessing the syntactic correctness of sentences and expressions can also be informative about the features of the construction of inner speech. Surveys, interviews and observations are methods actively used by representatives of a number of humanities, including psychologists, linguists and psycholinguists. For example, we used in-depth interviews to study the behavior of young people related to the Internet. The subjects were given a questionnaire consisting of 14 questions, the answers to which were supposed to provide both quantitative and qualitative characteristics. For example, a question containing evaluative information assumed a quantitative characteristic of the object: "From what sources do you get information? Determine its share in %". At the same time, the participants gave detailed comments to such questions as: "How often do you watch TV? What channels and what programs do you watch? If you never watch, then why?" The survey was conducted in written form, but we managed to talk to some informants. During an oral interview with participants, the interviewer has good opportunities to obtain detailed, comprehensive comments and provide explanations. An example of an oral detailed interview (deep interview) is presented in our monograph on the study of electronic communication of young people.

Unlike traditional regression models that primarily focus on mean responses dependent on values of covariates, fractile regression delves into the fractile of a response based on these covariate values (Castro et al., 2023). As Novikova-Grund (2017) mentions, the connection with the fact that such most important elements in existential and humanistic psychotherapy of existence as "the picture of the world of man", "freedom", "loneliness," etc. are semantically uncertain concepts and therefore they are inaccessible for experiment checkup with the use of the traditional mathematical methods. Other type of research includes organizational research methods which include:

a) comparative method, b) longitudinal method, c) complex method.

In the comparative method, or "sectional" method, different groups of subjects (for example, normal or pathological speech, preschool and primary school children) or different aspects of speech activity (for example, perception, speech production, reading, writing, inner speech) are compared. The longitudinal method, or "longitudinal sectional" method, involves a long-term study of certain components of speech activity. Such studies last for 5 years and are most often aimed at studying the language acquisition of children of different ages. The complex research method combines the features of comparative and longitudinal methods - in this case, the features of speech development or speech perception of representatives of different age groups are compared simultaneously (in parallel) (Ananyev, 1977).

In the analysis of verbal attitudes, we used the method of observation and collection of language material containing erroneous speech acts recorded by us in the speech of bilinguals. In carrying out this psycholinguistic study, we relied on the teachings of D. N. Uznadze on verbal attitudes, according to which the beginning of speech is preceded by a certain holistic change of the subject, a special state that directs the speech mechanism and acts in the subject. The collected material on erroneous speech acts made in the comprehension of foreign language texts made it possible to compile a classification of errors and to interpret them in terms of the influence of various types of formed verbal attitudes. For example, the "relativity" expressed in the functioning of tense forms in the Russian language, independent of the type of syntactic structure, leads to semantic distortions in comprehension. Thus, the sentence "The forest became dark as on the darkest night" was translated by the students as "The forest became dark as on the darkest night". The correct option is: "as in the darkest night" (Potebnya, 1999).

When designing and conducting psychological experiments, the possibilities of linguistic and psychological methods are taken into account. Traditional linguistic methods are aimed at studying language as a system-structural formation, and text as a set of semiotic signs. Among purely linguistic methods, we can mention: structural-functional, descriptive, componential, stylistic, comparative-contrastive, descriptive and other types of analysis. And it is now generally accepted that "… knowledge of language makes it possible to understand human consciousness" (Belyanin, 2003). Therefore, the use of semantic analysis can provide quite valid data from a psycholinguistic point of view. The connection between speech and consciousness is explained by A.R. Luria: "The word expresses things, the word emphasizes properties, actions, relationships. The word combines objects into known systems, in other words, it encodes our experience" (Luria, 1979). There are various psycholinguistic experiments based on the study of the semantic structure and functions of the word. For example, methods of word expansion and methods of determining meaning. Luria (1979) described a method for studying the awareness of the verbal composition of the language in the process of speech development. The study of the functioning of the categorization mechanism is based on the meaning of the word in categorical procedures.

In psycholinguistic experimental research, a method of semantic discrimination is needed. This assumption was made by Osgood et al. (1957), one of the founders of the discipline of psycholinguistics, and his colleagues. The goal of the experiment is to identify the semantic field of a certain concept. For example, informants are asked to define a concrete or abstract concept with several (at least three) parameters, usually using a seven-point rating scale. (Recently, surveys with the number of parameters from 5 to 9 have also been conducted, but these surveys are labor-intensive to complete and can give contradictory results). The choice of parameters and their properties is determined by the tasks determined in advance. For example, the concept of "strength" can be interpreted as dark - light, passive - active, slow - fast, chaotic - ordered, weak - strong, dangerous - safe, soft - hard, etc. It is recommended to conduct the assessment on a seven-point scale (from -3 to +3). We believe that the semantic differential method is especially effective in studying axiological entities, but the participation of a sufficient number of informants is necessary to obtain data suitable for interpretation.

The best-known psycholinguistic technique is the associative experiment. This is the most effective way to access information stored in the human mind, which is often called the "black box". "Unification (connection) is understood as a legal connection between two or more mental processes (emotions, ideas, thoughts, images, feelings, actions, etc.). Unification is expressed by the emergence of one of the processes, giving rise to another or other mental processes" (Spencer & Ziegen, 1998). In the first case, a distinction is made between correlation experiments; Subjects are asked to describe the reactions that arise in them in response to the presented stimulus word. In the course of a directed experiment, informants are limited in their choice of partners, and their total number determines the interpretation of the results obtained, i.e., the choice. Semantic connections and associations depend on the tasks set in the experiment and the semantic

content of the set of answers. Its content is analyzed and contains ready-made material for studying semantic relations (Ufimtseva et al., 2021).

In our work, we used relational experiments aimed at reconstructing the images of consciousness of representatives of various ethnolinguistic cultures. The scientific hypothesis was based on the fact that the acquired associations to a particular stimulus word and their choice are determined by the national and cultural features of the linguistic consciousness of informants. (The subjects of the study were native speakers of Russian, English, and bilingual people). All stimulus words were divided into six groups: lexical universals, words with the same meaning and national-cultural-specific connotation, "internationals" with a similar sound, hypernyms, and words denoting Russian and American realities. As expected, the greatest number of coincidences is usually recorded in responses to universal stimuli, and the greatest degree of national-cultural marking was recorded in responses to stimuli presented by real words.

CONCLUSIONS

Modern psycholinguistics presents itself as an interdisciplinary field that integrates the achievements of psychology and linguistics, thus enabling a deeper understanding of the interactions between language and cognition. Experiments are fundamental to this discipline, serving as essential tools for gathering relevant data that guide research. However, the uniqueness of each study—given the variety of subjects and objectives—highlights the need for a flexible and adaptable approach to methodology. Therefore, in addition to widely accepted methodologies such as the associative experiment and the semantic differential method, it is crucial to encourage the development of innovative experimental methods. The work of the prominent Russian psycholinguist I.N. Gorelov illustrates this need by proposing experimental techniques that examine the relationship between "sign and representation" and the characteristics of the synesthesia mechanism. This underscores that the repertoire of experimental methods in psycholinguistics is not static but constantly expanding, thus allowing researchers to creatively address complex and evolving scientific problems. Such methodological flexibility is vital to further advances in the understanding of cognitive processes related to language.

REFERENCES

Ananyev, B. G. (1977). ***On the problems of modern anthropology***. Nauka. http://elib.old.gnpbu.ru/text/ananyev_o-problemah-sovremennogo_1977/fs,1/

Belyanin, V. (2003). ***Psycholinguistics: A Manual for Graduate Students in Linguistics and Psychology***. Flinta, Nauka.

Boas, F. (2021). ***Anthropology and Modern Life***. Routledge & CRC Press. https://www.routledge.com/Anthropology-and-Modern-Life/Boas/p/book/9780367679910

Castro, C., Leiva, V., Lourenço-Gomes, M. do C., & Amorim, A. P. (2023). Advanced Mathematical Approaches in Psycholinguistic Data Analysis: A Methodological Insight. ***Fractal and Fractional***, ***7***(9), Article 9. https://doi.org/10.3390/fractalfract7090670

Duan, X., Zhou, X., Xiao, B., & Cai, Z. G. (2024). ***Unveiling Language Competence Neurons: A Psycholinguistic Approach to Model Interpretability*** (No. arXiv:2409.15827). arXiv. https://doi.org/10.48550/arXiv.2409.15827

Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). ***Neural Language Models as Psycholinguistic Subjects: Representations of Syntactic State*** (No. arXiv:1903.03260). arXiv. https://doi.org/10.48550/arXiv.1903.03260

Klyukanov, I. E. (2018). ***Message and forgetting***. Center for Humanitarian Initiatives.

Leontiev, A. (2003). ***Fundamentals of Psycholinguistics***. Smysl Publishing House.

Luria, A. R. (1979). ***Language and Consciousness*** (E. D. Khomskaya, Ed.). Publishing House of Moscow University.

Mead, G. H. (1962). ***Mind, Self and Society*** (C. W. Morris, Ed.; 2nd ed.). University of Chicago Press. http://tankona.free.fr/mead1934.pdf

Novikova-Grund, M. W. (2017). Psycho-semiotic approach in psycholinguistics analysis of verbal material as a means of recreation of the individual picture of the world of man. ***Journal of the Siberian Federal University. Humanities***, ***10***(3), 378–390.

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). ***The Measurement of Meaning***. University of Illinois Press. https://gwern.net/doc/psychology/1957-osgood-themeasurementofmeaning.pdf

Potebnya, A. A. (1999). ***Complete Works. Thought and Language***. Labyrinth.

Sacks, J. M., & Levy, S. (1950). The Sentence Completion Test. In L. E. Abt & L. Bellak (Eds.), ***Projective Psychology*** (pp. 357–402). Knopf.

Sajadi, A., Damevski, K., & Chatterjee, P. (2025). ***Psycholinguistic Analyses in Software Engineering Text: A Systematic Literature Review*** (No. arXiv:2503.05992). arXiv. https://doi.org/10.48550/arXiv.2503.05992

Sapir, E. (1949). ***Selected Writing of Edward Sapir in Language, Culture and Personality***. University of California Press. https://www.grammainstitute.com/wp-content/uploads/2024/03/Selected-Writings-of-Edward-Sapir-in-Language-Culture-and-Personality-PDFDrive-.pdf

Slobin, D. I. (1972). Psycholinguistics. ***Foundations of Language***, ***9***(2), 287–287.

Spencer, G., & Ziegen, T. (1998). ***Foundations of Psychology. Physiological Psychology in 14 Lectures***. AST-LTD Publishing House.

Ufimtseva, N. V., Vladimirovna, U. N., Balyasnikova, O. V., & Veniaminovna, B. O. (2021). National Identity and the Associative-Verbal Network: On a Hypothesis of Yu.N. Karaulov. ***RUDN Journal of Language Studies, Semiotics and Semantics***, ***12***(2), Article 2. https://doi.org/10.22363/2313-2299-2021-12-2-238-254

von Humboldt, W. (2000). ***Selected Works on Linguistics*** (G. V. Ramishvili, Ed.; A. V. Gulygina & V. A. Zvegintsev, Trans.). Progress Publishing House.

Vygotsky, L. S. (2001). ***Thinking and Speech. Psyche, Consciousness, Unconscious***. Labyrinth.

Whorf, B. L. (1956). ***Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf***. Technology Press of Massachusetts Institute of Technology. https://mitpress.mit.edu/9780262730068/language-thought-and-reality/