QUANTITATIVE METHODS IN ECONOMICS
Volume XXVI, No. 4, 2025, pp. 155 – 164

INVESTMENTS IN LANGUAGE CAPITAL: A NETWORK BASED ANALYSIS FROM THE HUMAN CAPITAL PERSPECTIVE

Robert Woźniak https://orcid.org/0009-0009-6020-5247
Mariola Chrzanowska: https://orcid.org/0000-0002-8743-7437
Institute of Economics and Finance
Warsaw University of Life Sciences – SGGW, Poland
e-mail: robert_woźniak@sggw.edu.pl; mariola_chrzanowska@sggw.edu.pl

Abstract: This study examines investments in linguistic capital as a component of human capital in a globalized economy. Addressing the lack of relational perspectives in existing research, the analysis is based on publicly available Duolingo data used to construct a directed and weighted network of language-learning flows. The results reveal a selective and hierarchical structure of investments, characterized by asymmetries, small-world properties, and a concentration of educational resources around languages with the highest expected economic returns. The findings indicate that linguistic capital is allocated through interconnected global structures shaped by network effects rather than independent individual choices.

Keywords: language capital; human capital; social network analysis; network effects; Duolingo

JEL classification: I21, J24, C88, Z13

INTRODUCTION

The increasing globalization and internationalization of labor markets give foreign language proficiency a growing economic importance [Jabłoński 2021]. Consequently, language competencies constitute a significant factor shaping the position of individuals and economies in the international financial system. In this approach, language competencies are treated as a component of human capital, influencing labor productivity, employment opportunities, and income levels.

https://doi.org/10.22630/MIBE.2025.26.4.14

According to human capital theory, educational decisions are investment-based and made based on expected returns [Becker 1964]. Learning foreign languages fits this pattern, as it involves incurring costs in time and money to obtain future economic benefits. Choosing a specific language can therefore be interpreted as a rational allocation of educational resources. The literature indicates that the return on investment in linguistic capital varies and depends on a given language's position in the global economic system [Bleakley and Chin 2004; Aldashev et al. 2009; Liwiński 2019]. Languages with an international function generate higher financial returns, favoring the concentration of educational investments in selected languages and leading to a hierarchization of linguistic capital. Empirical research to date has focused primarily on individual or national analyses. Less frequently considered is the fact that the value of linguistic competence also depends on the number of users and the structure of relationships between languages. In this sense, linguistic capital possesses the characteristics of a network good. Applying network analysis enables us to conceptualize investment in linguistic capital as a system of directed, weighted relationships among languages. This approach allows us to identify hierarchies, concentrations, and asymmetries in educational investments on a global scale. This article analyzes these relationships from the perspective of human capital theory, using social network analysis methods.

LANGUAGE COMPETENCES FROM THE HUMAN CAPITAL THEORY AND NETWORK ANALYSIS

In mainstream economics, linguistic competence is primarily analyzed within the framework of human capital theory. According to this concept, investments in education and skills increase productivity and income (Becker, 1964). In this perspective, foreign language proficiency constitutes a specific form of human capital, requiring time and financial investment while simultaneously generating future economic benefits. Decisions regarding investment in linguistic competence are rational and are made in response to market incentives and the expected rate of return (as confirmed by empirical research).

A significant portion of the literature focuses on analyzing the economic effects of foreign language proficiency, particularly wage premiums. Studies based on microeconomic data indicate that language proficiency significantly affects wages among both migrants and natives [Bleakley and Chin 2004; Aldashev et al. 2009]. In the European context, Liwiński [2019] demonstrates that the rate of return on investment in foreign language learning varies and depends on both the language type and the labor market. This confirms the existence of an economic hierarchy of language competences. This research by Hayfron [2001] complements this perspective. The author highlights the crucial role of formal language training as a prerequisite for achieving proficiency and entering the job market.

In parallel with this research, a growing body of literature emphasizes the relational and network nature of the value of language competences. The economic value of a given language depends not only on the individual proficiency of the user but also on the number and composition of other users. In this sense, a language can be treated as a network good, whose utility increases with the size of the community using a given standard [Katz and Shapiro 1986]. This mechanism favors the concentration of educational investments around languages with a strong international position. This perpetuates asymmetry and hierarchy in the global system of linguistic capital. The importance of structural context and linguistic focus is also emphasized by Chiswick and Miller [2002, 2007]. The authors point out that the returns to language competencies depend on the professional environment and the structure of the labor market (in which they are utilized). These results suggest that the economic value of linguistic capital is contextual and co-shaped by institutional and structural factors.

The rationale for the network approach is also reflected in research on language learning processes. Studies utilizing network analysis tools and complex systems theory have demonstrated that learning processes are dynamic and embedded in broader systems of relationships [Kiss 2020]. Analyses of data from learning platforms such as Duolingo reveal heterogeneous learning trajectories and strong aggregation effects. These results are not evident in individual analyses [Streeter et al. 2015; Settles et al. 2020]. Although these studies primarily focus on modeling teaching processes, they provide significant methodological arguments for treating investment in linguistic capital as a network phenomenon.

Despite the extensive literature on the economic consequences of foreign language proficiency, existing research has primarily focused on the individual or country level. Relatively rarely, however, has it analyzed the structure of language relationships as an element of the global human capital system. In particular, there is a lack of research combining human capital theory with formal network analysis to identify the hierarchy, concentration, and asymmetry of educational investments on a global scale. This research gap justifies the use of social network analysis methods in the study of linguistic capital.

Linguistic capital possesses both the characteristics of a network good and an economic good. The monetary value of a given linguistic competence depends not only on the individual characteristics of language users but also on the number and structure of users of a given language. The larger the population speaking a given language, the higher the potential return on investment in learning it. This mechanism leads to network effects, resulting in already widespread languages attracting further educational investments, strengthening their position in the system. As a consequence, the global market for linguistic capital adopts a hierarchical structure, with a clearly distinguished core and periphery.

THE NETWORK AS A MODEL FOR STRUCTURING INVESTMENT IN LANGUAGE CAPITAL

To capture the interdependencies between individuals' educational decisions, a graph theory formalism is employed. The global system of linguistic capital investments is modeled as a directed, weighted graph:

$$G = (V, E) \tag{1}$$

where V denotes the set of nodes representing languages, and E — a set of directed edges corresponding to educational investment flows. Each edge $e_{ij} \in E$ leads from language i to language j and has a weight w_{ij}, which reflects the intensity of investment in learning language j by users of language i.

Network Density:

Network density determines the degree of utilization of possible relationships between nodes and is defined as:

$$D = \frac{|E|}{|V|(|V|-1)} \tag{2}$$

In an economic context, this measure indicates the degree of integration of the linguistic capital market. Low density indicates selective educational investment, while higher density indicates greater diversification of decisions.

Characteristic path length:

The average length of the shortest paths in the network is given by the formula:

$$L = \frac{1}{|V|(|V|-1)} \sum_{i \neq j} d(i,j) \tag{3}$$

where $d(i,j)$ denotes the length of the shortest path between nodes i and j. A short average path length indicates high structural efficiency of the network and quick access to key resources of linguistic capital.

Network Diameter

The network diameter is defined as the maximum length of the shortest path between any two nodes:

$$\text{diam}(G) = \max_{i,j \in V} d\,(i,j) \tag{4}$$

This measure informs about the greatest structural investment distance in the system.

Clustering coefficient:

For node i, the clustering coefficient is given by the formula:

$$C_i = \frac{2e_i}{k_i(k_i-1)} \tag{5}$$

where k_i denotes the degree of a node, and e_i — the number of connections between its neighbors.

The average clustering coefficient in the network is:

$$C = \frac{1}{|V|}\sum_{i \in V} C_i \tag{6}$$

High clustering indicates the existence of local blocks of linguistic capital, such as regional or cultural groups.

Nodal measures – economic centrality of languages

Weighted input stage (in-strength)

$$s_i^{\text{in}} = \sum_j w_{ji} \tag{7}$$

This measure is interpreted as the demand for linguistic capital of a given language.
Weighted in-strength (Weighted out-strength)

$$s_i^{\text{out}} = \sum_j w_{ij} \tag{8}$$

It reflects the scale of outward investment, i.e. the adaptation pressure of users of a given language.
Betweenness centrality

$$BC_i = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{9}$$

where σ_{st} denotes the number of shortest paths between nodes *s* and *t,* and $\sigma_{st}(i)$ — the number of such paths passing through node *i*. This measure identifies languages that act as structural intermediaries in the global system of linguistic capital.

The network measures used enable a quantitative assessment of the structure of the global system of investment in linguistic capital. Structural measures such as density, average path length, and network diameter allow us to assess the degree of integration and the spatial scope of the linguistic capital market. Low network density indicates the concentration of investments around a limited number of languages. In contrast, a short average path length indicates high accessibility of key languages and the system's structural efficiency. Network diameter, in turn, refers to the maximum distance between languages and identifies potential barriers to accessing central linguistic capital resources.

Nodal measures enable us to identify the economic role of individual languages in the system. A high value for the weighted entry degree indicates strong demand for a given language and its high investment attractiveness. A high exit degree, on the other hand, reflects the adaptive pressure of language users to acquire competencies in other languages. Intermediation centrality identifies languages that act as structural intermediaries, connecting distinct network segments and facilitating the flow of investments between different areas of the linguistic capital system.

DATA SOURCE

The data were compiled from public information on the Duolingo platform (www.duolingo.com), including the number of users learning each language and the most common learning trends. Based on this data, a semi-synthetic, weighted, and directed language network was constructed, with nodes representing languages and edges representing estimated user flows (in millions).

PURPOSE OF THE STUDY

This study aims to analyze the patterns of foreign language selection among users of a global educational platform from a network perspective. It examines how languages function as sources and targets of educational investment, and how they relate to one another in the context of human capital theory and network effects. The study aims to identify central, peripheral, and intermediary languages in the global system of linguistic competences.

NETWORK ANALYSIS RESULTS

Table 1. Basic global statistics of the language network

Parameter	Value	Description
Nodes	18.00	Number of languages included in the analysis
Edges	63.00	Number of educational relations / investment flows between languages
Average number of neighbours	3.67	Average number of outgoing connections per node
Network diameter	2.00	Longest shortest path between any two languages
Network radius	1.00	Minimum of the maximum distances from any node to all others
Average path length	1.74	Mean length of the shortest paths between all pairs of languages
Clustering coefficient	0.49	Degree of local clustering of languages, indicating the presence of clusters (e.g. regional languages)
Density	0.21	Proportion of realised connections relative to the maximum possible number (integration of the language capital market)
Number of connected components	1.00	All languages form a single connected structure
Number of bidirectional pairs	30.00	Number of language pairs connected by mutual relations (e.g. two-way educational investments)

Source: own calculation

Economic interpretation of global parameters

Analysis of global network parameters allows us to assess the structure of investment in linguistic capital as a relational economic system. The studied network includes 18 languages and 63 directed investment relationships, indicating an ordered, non-random structure of connections. A network density of 0.206 suggests that only a limited portion of potential relationships between languages are realized. From the perspective of human capital theory, this indicates the selective nature of educational decisions and the concentration of investments around languages with the highest expected rate of return.

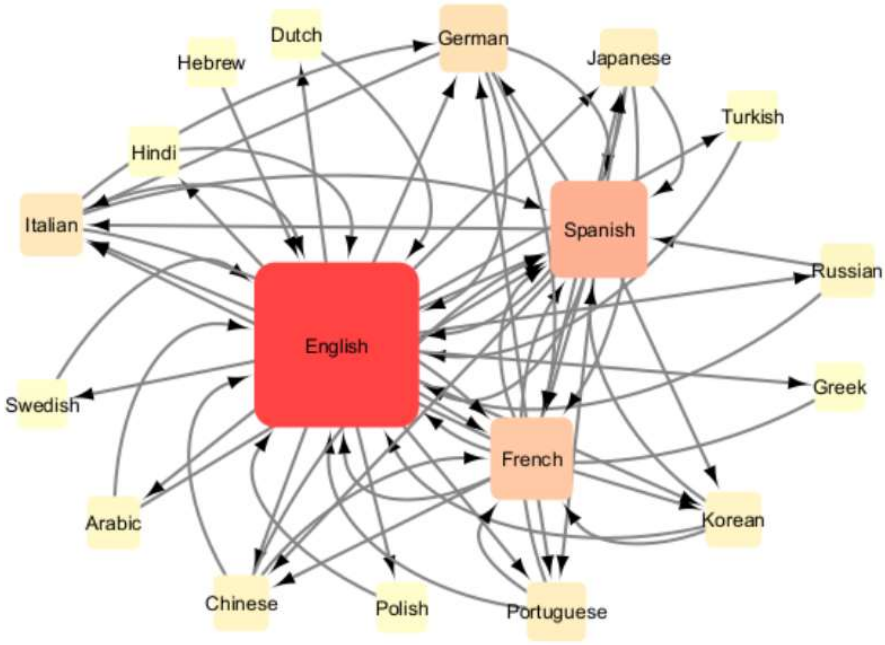
At the same time, the network forms a single coherent component, indicating the full integration of the global linguistic capital market and the absence of languages completely excluded from the system. The low average path length (1.738) and network diameter of 2 suggest a highly centralized structure, in which access to key linguistic capital resources is via a limited network core. This arrangement is characteristic of systems in which educational investments are channeled through languages with a dominant economic position. The relatively high clustering coefficient (0.498) also indicates the presence of local linguistic clusters, suggesting segmentation of the linguistic capital market stemming from cultural, geographic, or institutional determinants. The coexistence of short global paths and high clustering confirms the existence of a small-world structure. Furthermore, the presence of numerous bidirectional relations, with varying weights, reveals an investment asymmetry between languages, in which some act as net beneficiaries of linguistic capital, while others act as its sources.

Language network visualization

Figure 1 shows a directed, weighted language network of Duolingo users, generated in Cytoscape using a force-directed model. Nodes represent languages, and directed edges represent learning directions, from source to target language. The size of the nodes is proportional to the number of learners of a given language (in millions), while their color reflects the scale of this variable. The thickness of the edges corresponds to the intensity of user flows between languages.

The visualization confirms the existence of a highly centralized network structure, dominated by several high-centrality languages. English serves as the central node of the network, characterized by high in-degree and out-degree and high betweenness centrality. Languages such as Spanish and French occupy intermediate positions, acting as regional nodes that integrate flows of educational investment. This arrangement is consistent with both human capital theory and the network interpretation of investment processes.

Figure 1. Directed and weighted network of language capital investments based on Duolingo data



Source: authors' own elaboration

SUMMARY

This article aimed to analyze investments in linguistic capital from the perspective of human capital theory, using tools from social network analysis. Linguistic competence was treated as a specific form of human capital, requiring educational costs and generating diverse economic benefits, the scale of which depends on a given language's position in the global economic system. The use of graph formalism enabled a shift from the analysis of individual educational decisions to a structural description of the international linguistic capital market.

The obtained results indicate that the system of investments in linguistic capital is simultaneously highly integrated, selective, and hierarchical. The moderate network density confirms that decisions to learn foreign languages are not random but rather concentrate around a limited number of languages with the highest expected rate of return. A single coherent component and a very short average path length indicate a high level of global integration, within which all languages function within a single system of interdependencies. At the same time, a relatively high clustering coefficient indicates the presence of local linguistic clusters, suggesting

segmentation of the linguistic capital market driven by cultural, regional, or institutional factors. The combination of high global integration and local segmentation suggests a small-world structure, characteristic of many economic networks.

The presence of numerous bidirectional relations, with their simultaneous variation in weight, confirms the existence of asymmetric investment dependencies between languages and a structural imbalance in the global system of linguistic capital. These results are consistent with the assumptions of human capital theory, according to which educational investments are directed toward resources with the highest productivity, and with the concept of network effects, which leads to the concentration of capital around dominant linguistic standards.

The methodological contribution of the article lies in combining human capital theory with formal network analysis, enabling the identification of hierarchies, concentrations, and asymmetries in educational investments at the global level that are invisible in traditional analyses based on individual data or national aggregates. This approach allows for a quantitative understanding of the relational nature of linguistic capital as a network good.

However, it should be emphasized that the study is cross-sectional and relies on aggregated investment flows, limiting the ability to draw direct conclusions about individual behavior and temporal dynamics. Future research could expand the analysis to include a temporal dimension and link network parameters to external economic variables, such as the structure of international trade, migration, and levels of economic development. In summary, the results demonstrate that investments in linguistic capital are governed by network effects, leading to hierarchical and asymmetric patterns of educational resource allocation at the global level. These patterns can be examined and interpreted within the human capital framework using a network-based analytical perspective.

REFERENCES

Aldashev A., Gernandt J., Thomsen S. L. (2009) Language Usage, Participation, and Earnings: Evidence for foreiGners in West Germany with Multiple Sources of Selection. Labour Economics, 16(3), 330-341.

Becker G. S. (1964) Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. University of Chicago Press.

Bleakley H., Chin A. (2004) Language Skills and Earnings: Evidence from Childhood Immigrants. Review of Economics and Statistics, 86(2), 481-496.

CASE Center. (2023) showCASE 87: Language and Economic Inclusion. Retrieved from https://case-research.eu

Chiswick B. R., Miller P. W. (2002) Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle. Journal of Population Economics, 15(1), 31-57.

Chiswick B. R., Miller P. W. (2007) The Economics of Language: International Analyses. Routledge.

Chiswick B. R., Lee Y. L., Miller P. W. (2006) Immigrants' Language Skills and Visa Category. IZA Discussion Paper No. 3568.

Hayfron J. E. (2001) Language Training, Language Proficiency and Earnings of Immigrants in Norway. Applied Economics, 33(15), 1971-1979

Jabłoński Ł. (2021) Ewolucja podejść do kapitału ludzkiego w naukach ekonomicznych [Evolution of Approaches to Human Capital in Economic Sciences]. Gospodarka Narodowa, 2(306), 91-120.

Katz M. L., Shapiro C. (1986) Technology Adoption in the Presence of Network Externalities. The Journal of Political Economy, 94(4), 822-841.

Kiss T. (2020) Network Analysis for Modeling Complex Systems in SLA Research. Studies in Second Language Learning and Teaching, 10(2), 335-358.

Liwiński J. (2019) Returns to Foreign Language Skills in a Post-Transition Economy: Evidence from Poland. Education Economics, 27(3), 225-242.

Settles B., Meeder B. (2020) A Trainable Model for Learner Sourcing. Transactions of the Association for Computational Linguistics, 8, 1-14.

Streeter M., Settles B., Boyd A. (2015) Crowdsourcing Multiple-Choice Questions for Natural Language Grammar. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 13-22.