

First steps into scientific research



Teacher: Jelena Budimirović

Teacher: Maja Đukanović

Teacher: Jelena Gledić

Teacher: Maja Miličević Petrović

Teacher: Novella Tedesco


This 3 ECTS learning block (with a possibility of adding a 1 ECTS project) introduces the basic concepts of scientific research, outlining the main approaches in science in general and the key steps in the research process as applied to language. After completing the learning block, the student will be able to:

- approach a subject from a scientific perspective
- evaluate the methodology of scientific articles
- relate a research question to a scientific theory and a research experiment
- design and setup a simple research proposal and plan

Benefit from existing materials - Movetia online course *Revisiting research training in linguistics: theory, logic, method*

Definitions with examples that can be used individually as needed

Based on how their levels are expressed, variables can be divided into:



categorical (also called qualitative) → these are variables whose values are expressed as categories

numerical (quantitative) → these are variables whose values are expressed as numbers

In some cases, the decision between these two types is imposed by the nature of a variable (e.g., sex or part of speech can only be expressed as categories). But in many cases the decision is up to the researcher.

0:00 / 2:35

2.2. Data sampling

In scientific research, most of the time we try to reach conclusions that apply to a **population**, but we can only study a **sample** taken from that population.

A population is composed of all members of some group that share some properties. Examples of populations are all citizens of a country, or all students of a department. In language studies, populations can consist of **all speakers** of a language or a combination of languages (e.g., native speakers of German, bilingual speakers of German and Polish, speakers of Polish as a second language), or of **all texts** in a given language, within a given variety, genre, and similar. For texts, the population is the **entire language**, variety, or genre (e.g., Hungarian, or spoken Sinhala). The specific group the term "population" refers to is determined relatively, depending on the research topic.

Populations are extremely rarely accessible to researchers in their entirety. This is typically because they are too large (can you imagine a study that tests ALL native speakers of Turkish?). Sometimes populations don't even have a clear limit – think of what you would need to study if you wanted to explore the whole of the Bulgarian language (could it be everything every native speaker has ever said?).

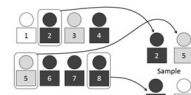
This is why in individual studies, we mostly rely on samples, which are composed of representatives of the population of interest. When you conduct a study with speakers of a language, your **participant group(s)** will be your sample. In the case of texts, typical samples are **electronic corpora**.

A common component of data analysis concerns making conclusions about populations based on data from samples. This is what **statistical analysis** is largely about, and in particular **statistical tests**.

A good sample needs to be **representative** of the population, i.e. to reflect its properties closely. For example, if a population of second language learners of Macedonian is composed of approximately 50% women and 50% men, a good sample will reflect this (rather than having 80% women and 20% men).

The most representative kind of samples are **random samples**. The members of a random sample were randomly selected from the population -> each member of the population had an equal initial chance of being selected.

If you assume that you have a numbered list of all members of the population, random sampling can be based on a list of random numbers (ideally generated by a computer algorithm, not by the researcher). Or you can decide to take every *n*-th member of the population. [\[references for images to be added, or images replaced\]](#)



Practical exercises, for example:

Making a research plan

Select one of the settings listed below and make a draft research plan for how to study it.

- You are a researcher who wants to explore the most common **citation markers** in informal conversations of young people – in other words, how they introduce the material they quote. They could, for example, use linguistic means, or paralinguistic means such as gestures.
- Foreign urban **microtoponyms** (names of squares, streets, important buildings, etc.) are very often mentioned in different types of media. They can be reported in the original form, transcribed, calqued, translated. How would you research the way they are used in your native language?
- You are exploring the **degree of motivation for learning Slovenian** among the members of the Slovenian minority in Serbia and those who do not belong to it. You conduct research with students within the Society of Slovenians in Belgrade, who all have Slovenian roots, and with students of the Slovenian language at the University of Belgrade, who do not have Slovenian origins.

In the research plan, regardless of the topic, keep in mind the following questions:

1. Where would you look for ideas about a theoretical framework to apply to the study?
2. Would you go for a qualitative or a quantitative study?
3. Would you go for observational or experimental research?
4. Would your study include a baseline condition?
5. Would you narrow the topic down, and if so, how?
6. What would be your specific research questions?
7. Would you have a hypothesis?
8. What kind of research design would you implement?

Critically thinking about the relevance of cultural context in scientific research

Consider two books:



A Dictionary of the English Language, compiled by Samuel Johnson

Image: Samuel Johnson, Public domain, via Wikimedia Commons



The Kangxi Dictionary of the Chinese language, commissioned by Emperor Kangxi and compiled by tens of scholars

Image: Malcolm I'Anson, Public domain, via Wikimedia Commons

Both books are influential dictionaries first compiled in the 18th century.

Find relevant sources to learn more about these dictionaries, and then consider and discuss the following questions:

1. What are the differences and similarities in their structure?
2. How might have the differences between the languages influenced the way the dictionaries were compiled?
3. Based on the structure of the dictionaries, what would you assess was their intended use?
4. What kind of knowledge and skills would a person need to access the information provided in each of the two dictionaries?
5. Considering your answers to the previous questions, compare how literacy was understood in 18th-century England and China.

An **additional activity** is to explore the digital versions of these dictionaries:

<https://johnsonsdictionaryonline.com/index.php>

<https://ctext.org/kangxi-zidian>

We suggest you look up the same words in both versions, for example:

Earth 地

Water 水