

Why do children's words matter?

The number of words children know is one of the strongest indicators of their language development. Psychologists and speech therapists around the world use vocabulary tests to determine whether a child is developing typically or may need additional assessment or support. Problems arise, however, when we try to compare the results of children who speak different languages. Is a vocabulary test in Polish as difficult as its Turkish, Finnish, or Spanish counterpart? Can bilingual children be assessed fairly when their two languages differ substantially? And if we use two tests, can we reliably determine which of a child's languages is stronger?

To answer such questions, researchers developed the Cross-Linguistic Lexical Tasks (CLTs) — a set of vocabulary tasks available today in more than 40 languages. All versions follow the same design principles: they have the same structure, similar questions, and the same number of pictures. Because CLTs are constructed according to the same rules in every language, they *should* also have similar levels of difficulty. In practice, however, no one has yet tested whether these tasks truly function in the same way across languages. This means that researchers and clinicians may be drawing conclusions based on an assumption that has never been sufficiently verified.

Project objective

The goal of this project is to conduct the world's first comprehensive evaluation of whether CLTs are genuinely comparable across languages, and to develop new, data-driven norms for bilingual children. The study will analyse large collections of data from multiple countries. We plan to include both published datasets and previously unavailable materials — for example, data from master's theses or from languages that are less frequently represented in scientific journals.

How will we do this?

The project consists of three studies:

- 1. Comparing test difficulty across languages.**

We will check whether children speaking different languages achieve similar results when they are of the same age and share similar socio-economic backgrounds. This will allow us to determine whether CLTs are equally difficult across languages.

- 2. Identifying which properties of words make them easy or difficult.**

We will examine whether the same characteristics of words affect difficulty across languages. For example:

- Does word frequency always influence how easy a word is?
- Does it matter at what age a word is typically learned?
- Or how easily its meaning can be imagined?

- 3. Creating new norms for bilingual children.**

We will determine how a child's vocabulary knowledge can be predicted based on age, the amount of exposure to each language, and the degree of similarity between the child's two languages.

Why does this matter?

Bilingual children are sometimes incorrectly diagnosed as having delayed language development. This often happens because we lack tools that fairly assess both of their languages. The results of this project will change this in two ways. First, they will show whether different language versions of the CLT can be meaningfully compared. Second, they will provide norms for bilingual children that can be used for *any* pair of languages.

Expected outcome

The project will create the largest to date multilingual database on vocabulary development in preschool-aged children. Its results will support researchers, clinicians, and practitioners worldwide. Ultimately, the project will contribute to fairer assessment of language development and facilitate more accurate diagnosis in bilingual children — a rapidly growing group in today's societies.