Letter of intent

File number 175.2021.049
Grant 2021/2022

Applicant
Title Corpus Base for Acquired Language Disorders in Dutch (COBALD) Fostering research into language disorders and the development of inclusive language and speech technology

Abstract
NWO-domein: Sociale en Geesteswetenschappen (SGW)

Discipline code(s):
30.40.00 Psycholinguistics and neurolinguistics, 37.40.00 Databases for humanities, 16.60.00 Artificial intelligence, expert systems.

Indication of the proposed research group:
Radboud University, Faculty of Arts:

• dr. Marina Ruiter (Centre for Language Studies)
• dr. Henk van den Heuvel (Centre for Language and Speech Technology)
Radboud University, Faculty of Social Sciences:
• dr. Vitória Piai (Donders Centre for Cognition, and also affiliated to Donders Centre for Medical Neuroscience, Radboudumc)
• dr. Willemijn Doedens (Donders Centre for Cognition)
MPI The Language Archive:
• Paul Trilsbeek (The Language Archive, Max Planck Institute for Psycholinguistics)

Connection with Groups(s) on the GWI Roadmap: Sociale & Geesteswetenschappen (Humanities/ Social sciences) If you do not see a connection with the GWI Roadmap, please indicate this explicitly: N/A.

Organisation responsible for the application

Confirm letter of intent
With submitting this form via ISAAC I declare to have filled in this form completely and truthfully.

✓

Motivation for the requested investment:

Today, one in four Dutch speakers lives with brain damage. The number of people with chronic brain damage is expected to grow exponentially. For many, accompanying acquired language disorders (ALD) significantly affect their independence and well-being. More research is critical to enable Dutch ALD speakers to fully participate in society, today and in the future. Large-scale data collections (corpus henceforth) are critical for language research and innovation in assistive technology, yet they are challenging and costly to collect and share. The interdisciplinary COBALD team meets these challenges by establishing the first accessible, longitudinal database of audio/video recordings of Dutch speakers with prevalent ALD (e.g. post-stroke aphasia and dementia). Paramount features of the corpus are its diversity in elicitation tasks, capturing relevant language variability and its longitudinal dimension, capturing temporal characteristics of the disorders. COBALD will be enriched with derived acoustic and language models intended for personalised automatic speech recognition (ASR). COBALD will enable (comparative and cross-linguistic) research in ALD and the development of speech technology tools for Dutch. The COBALD corpus will be made available to researchers and clinicians through the Language Archive (TLA), a CLARIN data centre as part of the CLARIAH infrastructure. It will be aligned with and incorporated into international databases such as TalkBank to maximise its exploitability for (inter)national research. COBALD will be fully GDPR-compliant.