



Press Release

Harald Baayen receives ERC Advanced Grant for groundbreaking language analysis project

University of Tübingen linguist will use the nearly 2.5 m euros to decouple language theory from the letter= sound constraint

Dr. Karl Guido Rijkhoek
Director

Janna Eberhardt
Research reporter

Phone +49 7071 29-76788
+49 7071 29-77853

Fax +49 7071 29-5566
karl.rijkhoek[at]uni-tuebingen.de
janna.eberhardt[at]uni-tuebingen.de

www.uni-tuebingen.de/aktuell

Tübingen, 12.04.2017



Harald Baayen Photo: Friedhelm Albrecht

Professor Harald Baayen of the University of Tübingen's Institute of Linguistics has obtained a nearly 2.5 million euro Advanced Grant from the European Research Council (ERC) for a radical new project in language analysis. The five-year project is called **WIDE** and aims to deepen our understanding of how we produce and understand

words in everyday speech.

Words in day-to-day conversational speech may differ substantially from how they appear in writing: German "würden" is often pronounced as "wün," Dutch "natuurlijk" ('naturally') can reduce to "tk", and Mandarin 要不然(jao pu zan, 'otherwise') to "ui." Current theories assume that the sound waves that reach our ears are reduced to sequences of abstract sound units, much like the sequences of letters that make up written words. However, how to align highly reduced forms such as "wün", "tk" and "ui" with their full unreduced variants, the supposed gatekeepers to meaning, is an unsolved computational problem.

The WIDE project makes the radical proposal to eliminate letter-like sound units altogether, and instead to zoom in on the rich details of the speech signal itself. Given tens of thousands of smart features representing the richness of the speech signal, it is anticipated that artificial neural networks can learn, by trial and error, to identify which meanings are conveyed. Previous research funded by the Alexander von Humboldt foundation allowed Baayen to provide a first proof of concept.

In the WIDE project, this approach will be developed further and extended from German to other languages, including Mandarin Chinese (a tone language) and Estonian (a complex language with 28 to 40 different forms for a given noun). The WIDE project also targets a computational model without sound units for the articulation of words in speech production.

The project's name, "WIDE", highlights a second aspect in which this project makes a radical departure from current trends in linguistics and natural language processing. Instead of making use of deep learning networks, the project focuses on the potential of 'wide' two-layer networks with tens of thousands of input and output units.

Rolf Harald Baayen, born in 1958 in the US, came to Tübingen in 2011 as a prestigious Alexander von Humboldt Professor. He holds the Professorship of Quantitative Linguistics, which was specially created for him. Previously, Baayen conducted his research at the University of Alberta in Edmonton, Canada. He is considered one of the world's foremost experts in the field of quantitative linguistics and vocabulary research. He is a pioneer in computer-assisted and empirical language research as well as in psycholinguistics; Baayen has made fundamental contributions to our understanding of human language and the role of memory in language processing.

The European Research Council uses Advanced Grants to support established, active researchers with an outstanding academic record. Advanced Grants can go to innovative projects in any discipline. Each project sponsored receives up to 2.5 million euros for up to five years.

Contact:

Professor Harald Baayen
University of Tübingen
Humanities
Institute of Linguistics
Phone +49 7071 29-73117
harald.baayen[at]uni-tuebingen.de