



# Laudētur Peter!

Building the Intelligent Infrastructure.  
On the Conferral of **Peter Wittenburg's**  
Honorary Doctorate

University of Tübingen  
Nov. 8, 2013

John Nerbonne



# Peter Wittenburg

- > Rising importance of scholarly infrastructure
- > Engineering for Linguistics and for the Humanities
- > Data-intensive science and scholarship
- > Cultural preservation
- > Passionate leadership



# Scholarly infrastructure

- › Linguistics, other humanities, in principle data-rich studies
  - 200 wd/min, 2.5 hr. verbal interaction/day =>  $10^7$  words/year
    - + intonations, gestures, non-verbal comm.
  - Google books estimates 130 million books published =>  $10^{13}$  words
  - Visual art, everyday artefacts, fashion, non-material culture ...
- › Not easily accessible!



# Studying lots of data

- > Recording, storing
  - High fidelity needed, lots of storage
- > Accessing
  - Not enough to "google"
    - Non-textual data: acoustics, images, video
    - Text before spelling normalization often shows 100's of variants *Tübingen, Tüübingen, Tüpbingen,*
  - Special programs needed
- > Selection: Annotation & Filtering
  - Scientific questions often very specific



# Engineering needed!

- › For storage, organization, annotation, access, coordination, compressing, analysis of data
  - Language and humanities have mostly “lifted” on growing demand from industry
- › Engineering **vision** needed!
  - Where are scientific needs *sui generis*, so that “lifting” is a poor strategy?
  - Where are there special cultural needs that won't be financed w.o. special attention?



# In principio erat ...

- > 1974 M.Sc. Electrical Engineering, TU
- > 1976 Head of new technical group, MPI Psycholinguistics (Nijmegen)
- > Install, maintain PDP 11's (64K)
  - Compare smartphones w. 1 GB!
  - Psycholinguistic exp., data storage
- > Pim Levelt: "then the roles switched"
  - Researchers began asking Pim how to take advantage of technology





# Projects!

- › MPI soon world leader psycholinguistics
  - Pioneer in technical psycholinguistics
    - Phonetic analysis
    - ERP/EEG (1980's)
- › CELEX (1GB, air-cooled memory) 80's
  - >3,500 citations (Baayen et al.)





# Corpora (text collections)

- > CHILDES (from early 90's on)
- > EUDICO (European distributed corpora, late 90's)
  - Distributed access
- > Dutch spoken language corpus (CGN,  $10^7$  words)
- > ISLE (Intl. Standards for Lang. Engineering)  
metadata initiative (2001)
  - Note early interest!
- > INTERA, ECHO (cultural heritage)
  
- > Lots on "his own initiative" (MPI director)



# Multimedia

- › Levinson => MPI, 1991
- › More than just text
- › Multimedia annotation
  - ELAN
- › Increasing interest for endangered languages
- › DoBeS (VW Stiftung), Virtual Language Observatory
- › The Language Archive (KNAW, BBAW, MPI Nijmegen)
- › Increasingly archiving linguistic material from all over the world





# Challenges wrt Endangered Languages

- > Not just technical!
- > Political, legal and ethical problems
- > Agreements needed
- > Funai, Brazil





# A brief Dutch lesson

- > English 'boss' is a Dutch borrowing (*baas*), and there are lots of derogatory terms for big shots, show offs *Bonzen, kapsones, kakkineus, dik doen, naast z'n schoenen lopen, ...*
- > *bobo* < *bonzen*, but also 'nitwit' (Surinam)
- > First use in Dutch dated to 1988 in reference to front office in football



# Ruud Gullit



> Known locally, too





# Peter W. as *bobo*

- › Active in many projects, boards, organizations
  - CLARIN (ERIC)
  - EUDAT
  - Research Data Alliance
- › ... *den neuen Status mit Fassung zu tragen*



Cafe Fitzcaraldo, Peru



## *... and the man?*

- > Dedicated and hard-working
- > Very popular among colleagues
- > Famously modest "geen gedoe rond dat erendocoraat"
- > Maintains some interest outside of work ...



# Football, for example ...

- > Never totally forgotten!
- > Improved Brazilian/Peruvian connections ...
- > PeWi in Rio!





# Laudētur Peter

> In praise of Peter!

- With thanks to Pim Levelt, Daan Broeder and Harald Baayen for ideas and photos!