Dupont Summit 2013

:: Pressing Issues, Economic Realities

December 6, 2013 ✳ Historic Whittemore House, Washington, DC

The Policy Studies Organization

Speaker

Andras Kornai - *Budapest Institute of Technology, Hungarian Academy of Sciences*

"Language Death in the Digital Age"

Of the approximately 7,000 languages spoken today, some 2,500 are generally considered endangered. We collected data from many sources and built a statistical model that makes clear that this consensus figure vastly underestimates the danger of digital language death. In fact over 95% of the worlds languages are already digitally dead, and only a few hundred can ascend to the digital realm. We present evidence of a massive die-off caused by the digital divide, not some future event, but one that has already taken place. To quote from the paper summarizing the work, this is

"the final act of the Neolithic Revolution, with the urban agriculturalists moving on to a different, digital plane of existence leaving the hunter-gatherers and nomad pastoralists behind."

This die-off has very broad `backward' policy implications for dealing with languages in the Third World and elsewhere, and also `forward' implications for dealing with the lucky 4-5% of languages that have made the transition to the digital world. The panel would involve experts from academia (both linguists and economists), industry, government, and from the digital domains where the remaining languages fight for survival, Wikipedia in particular.

Biography

András Kornai earned his mathematics degree in 1983 from Eötvös Loránd University, Budapest where his advisor was Miklós Ajtai. He earned his linguistics degree in 1991 from Stanford University, where his advisor was Paul Kiparsky. He is adjunct professor at the Budapest Institute of Technology, and Senior Scientific Advisor at the Computer and Automation Research Institute of the Hungarian Academy of Sciences. His research interests include all mathematical aspects of natural language processing, speech recognition, and OCR.