

# What do we Evaluate when we Evaluate Word Spaces, and what \*Should\* we Evaluate in Order to Evaluate them?

Magnus Sahlgren

SICS, Swedish Institute of Computer Science

Kista, Sweden

[mange@sics.se](mailto:mange@sics.se)

:

## **Abstract**

It has become increasingly popular to use vector spaces generated from distributional statistics as a representation of word meaning. The underlying structuralistic hypothesis is that the distributional profiles of words are symptomatic for their semantic content, and that a geometric representation of these profiles is cognitively plausible. Both the distributional hypothesis of word meaning and the geometric representational scheme have proven their mettle in many different experimental settings, with a wealth of publications reporting exceptional results and arguing for the viability of the approach. It is, to say the least, an active area of research. Even so, neither the reliability nor the validity of the evaluation methodologies have received much attention. This is remarkable since it is anything but obvious how to evaluate models of word meaning. To make matters even worse, it is not even obvious what "meaning" means. Conceptual opaqueness is all too often neglected in favor of experimentalism within the field of computational linguistics. However, when neither the conceptual nor the experimental basis is scientifically sound, one may seriously question the validity of the research. This paper does not attempt to solve the conceptual riddle about the meaning of "meaning". Neither does it problematize the distributional hypothesis nor the geometric representation. Rather, it problematizes the evaluation methodology. The following question constitutes the central theme of the discussion: How can we scientifically determine whether the geometric representations really do contain information about word meaning?