Monosemanticity spectrum (decoder.layer3)