Sequence to sequence learning: one-size fits all?