1. 基于语义距离的AUT\_originality计算
   1. 输入：被试+在每个原始物体上的+所有回答
   2. 预处理：
      1. 去除：去除所有回答中的一些“无关词“（如”的“，具体去什么由你决定）
      2. 替换：做一个替换列表，把一些常见的有问题的回答替换成相同意义的回答
      3. 句式调整（是否需要待定）：是不是需要把所有回答都调成 XX用来XX？这样效果会更好还是更差。也许不需要调，保留原始回答。到时候都试试。
      4. 其他……待添加
   3. 算每个回答与对应原始物体的语义距离（Bert + GloVe + word2vec + LSA，重要性依次递减，先找考前的模型做）
   4. 算出每个被试在每个物体上，所有回答到原始物体的平均语义距离，例如score\_brick
   5. 把每个被试的所有物体的score简单算个平均，作为每个被试的originality score。输出一个table，被试+originality score，就ok\
2. 语义聚合
   1. 输入：每个物体的所有回答（所有被试的所有回答）
   2. 预处理：
      1. 去除：去除所有回答中的一些“无关词“（如”的“，具体去什么由你决定）
      2. 替换：做一个替换列表，把一些常见的有问题的回答替换成相同意义的回答
      3. 句式调整（是否需要待定）：是不是需要把所有回答都调成 XX用来XX？这样效果会更好还是更差。也许不需要调，保留原始回答。到时候都试试。
      4. 其他……待添加
   3. 对这些回答做hierarchical clustering
   4. 找到最优cluster总数
      1. 最优的标准：还没太搞清楚，[The development of a valid and usable creativity test requires big-data and psychometrics (modelingcreativity.org)](https://modelingcreativity.org/blog/wp-content/uploads/2020/07/ABBAS_report_200711_final.pdf)，这篇文章用的是所有cluster内的originality scores的方差的平均值最小（感觉不应该用平均值，应该用总和，不然岂不是cluster越多就越容易均值小）。
   5. 然后我们一起给每个cluster起个名字，看看合不合理，然后把所有回答都标记上他们所属的cluster
   6. 然后计算每个被试所产生的一系列回答，一共涉及几个cluster（flexibility），在不同cluster间switch了几次（例如，第一个回答计第1次switch，第二个回答和第一个同属同一个cluster，则不计分，第三个回答与第二个回答的cluster不同，就再计1次switch）。最后得到两个指标，每个被试在每个物体上的flexibility和switch。
   7. 最后计算被试在所有物体上的flexibility和switch的均值