定量文学分析基本学习

很多人诟病这种方法，声称主要适用于context检验（e.g.文学史）而不是text本身的检验，对于text本身没啥意义（主要是美学感受学方面的concern）

但是对于e.g. 【作家的用词风格、造句模式、修辞习惯】 等有用，**但是做到啥程度了呢？（我还没调研）**

* 有些学者e.g.斯坦福的Moretti做过图谱分析，i.e.把人物视作图顶点，人物关系为图的边
  + **这个蛮有意思，甚至有clustering等应用和insights，尤其是可视化的insights**

Data是一大困难（e.g.跨库检索），目前主要是两大思路：

* Context, e.g. from platforms like Amazon/book reviews
* Long text, focus on text itself

用定量手段做情感分析的困难：

* 有些词汇潜藏在被检索的词汇范围之外而无法辨识（这个现在应该攻克了吧？我猜应该已经有工具/方法可以自动发现关键词汇了即使不是高频词也有重要词汇吧？）
* 价值判断
  + 这个可以调研一下，我猜特征提取可以给我们提供一些insights，什么是重要的特征，能反映价值观
  + 但反映价值观有什么用处呢？prediction比较有用吧，预测好的作品可以帮助出版业和作者；但天才的作品无法预测；但我们能做到前者已经很好了