Loup Ventures 风险资本的创始人蒙斯特（Gene Munster）准备了 800 个问题，对这几大智能语音助手进行了实际测试，最终评比结果涉及到两个关键指标：

1. 语音助手能否识别你提出的问题？

2. 语音助手能否理解问题，并提供正确的答案？

而在问题种类上，为了能尽可能涵盖多个领域和使用场景，蒙斯特将 800 个问题划分为 5 个部分：

1. 地点类，比如「离我最近的咖啡馆在哪里？」

2. 消费类，如「你能帮我订几筒卷纸吗？」

3. 导航类，如「怎么去附近的公交车站？」

4. 资讯类，如「今天天气怎么样？」

5. 指令类，如「请帮我打一个电话给 XXX」

问题：「请问修指甲要多少钱？」

Alexa 答：「我们为你搜索到 Beurer Electric Manicure & Pedicure Kit（一个修甲套装），亚马逊上的售价为 59 美元，请问你需要吗？」

Google Assistant 答：「相对来说，基础修甲服务为 20 美元，但一些包含光疗、凝胶等特殊服务的则为 20-50 美元不等，具体价格取决于美甲沙龙的情况。」

用户更希望实现的，是语音助手能够真正理解问题背后所隐藏的多重含义。

比如当你出门在外，提出「我的下一个待办事项是？」这个问题，除了单纯陈述事件外，提问者更愿意听到的回答可能是「你的下一次事项在 XX 分钟后，地点在 XXX，开车需要 XX 分钟，坐公交需要 XX 分钟，我们会给你导航方向」这种结合具体场景的应答方式。

我们注意到：正确的单词切分取决于对文本语义的正确理解，而单词切分又是理解语言的最初的一道工序。这样的一个“鸡生蛋、蛋生鸡”的问题自然成了（中文）自然语言处理的第一条拦路虎。

其他级别的语言单位也存在着各种歧义问题。例如在短语级别上，“进口彩电”可以理解为动宾关系（从国外进口了一批彩电），也可以理解为偏正关系（从国外进口的彩电）。又如在句子级别上，“做手术的是她的父亲”可以理解为她父亲生病了需要做手术，也可以理解为她父亲是医生，帮别人做手术。总之，同样一个单词、短语或者句子有多种可能的理解，表示多种可能的语义。如果不能解决好各级语言单位的歧义问题，我们就无法正确理解语言要表达的意思。

另外一个方面，消除歧义所需要的知识在获取、表达以及运用上存在困难。由于语言处理的复杂性，合适的语言处理方法和模型难以设计。

例如上下文知识的获取问题。在试图理解一句话的时候，即使不存在歧义问题，我们也往往需要考虑上下文的影响。所谓的“上下文”指的是当前所说这句话所处的语言环境，例如说话人所处的环境，或者是这句话的前几句话或者后几句话，等等。假如当前这句话中存在指代词的时候，我们需要通过这句话前面的句子来推断这个指代词是指的什么。我们以“小明欺负小亮，因此我批评了他”为例。在其中的第二句话中的“他”是指代“小明”还是“小亮”呢？要正确理解这句话，我们就要理解上句话“小明欺负小亮”意味着“小明”做得不对，因此第二句中的“他”应当指代的是“小明”。由于上下文对于当前句子的暗示形式是多种多样的，因此如何考虑上下文影响问题是自然语言处理中的主要困难之一。

再如背景知识问题。 正确理解人类语言还要有足够的背景知识。举一个简单的例子，在机器翻译研究的初期，人们经常举一个例子来说明机器翻译任务的艰巨性。在英语中“The spirit is willing but the flesh is weak.”，意思是“[心有余而力不足](https://www.baidu.com/s?wd=%E5%BF%83%E6%9C%89%E4%BD%99%E8%80%8C%E5%8A%9B%E4%B8%8D%E8%B6%B3&tn=24004469_oem_dg&rsv_dl=gh_pl_sl_csd" \t "https://blog.csdn.net/R1uNW1W/article/details/_blank)”。但是当时的某个机器翻译系统将这句英文翻译到俄语，然后再翻译回英语的时候，却变成了“The Voltka is strong but the meat is rotten.”，意思是“伏特加酒是浓的，但肉却腐烂了”。从字面意义上看，“spirit”（烈性酒）与“Voltka”（伏特加）对译似无问题，而“flesh”和“meat”也都有肉的意思。那么这两句话在意义上为什么会南辕北辙呢？关键的问题就在于在翻译的过程中，机器翻译系统对于英语成语并无了解，仅仅是从字面上进行翻译，结果自然[失之毫厘，差之千里](https://www.baidu.com/s?wd=%E5%A4%B1%E4%B9%8B%E6%AF%AB%E5%8E%98%EF%BC%8C%E5%B7%AE%E4%B9%8B%E5%8D%83%E9%87%8C&tn=24004469_oem_dg&rsv_dl=gh_pl_sl_csd" \t "https://blog.csdn.net/R1uNW1W/article/details/_blank)。

从上面的两个方面的主要困难，我们看到自然语言处理这个难题的根源就是人类语言的复杂性和语言描述的外部世界的复杂性。人类语言承担着人类表达情感、交流思想、传播知识等重要功能，因此需要具备强大的灵活性和表达能力，而理解语言所需要的知识又是无止境的。那么目前人们是如何尝试进行自然语言处理的呢？

例如，“你能把盐递过来吗”，一个好的回答应当是把盐递过去；在大多数上下文环境中，“能”将是糟糕的回答，虽说回答“不”或者“太远了我拿不到”也是可以接受的。再者，如果一门课程去年没开设，对于提问“这门课程去年有多少学生没通过？”回答“去年没开这门课”要比回答“没人没通过”好。

区分这三个句子是否是同义？张三把苹果吃了；苹果被张三吃了；张三吃了苹果。

这里面很复杂，绝对不是靠字面做出来的，涉及到深度语义，然后还需要进一步比较，但理解出来之后，就是匹配关系了。另外，这三句话的意思也是有差别的，例如，第一句话侧重点在吃的动作，第二句话的侧重点在苹果，第三句话在整句话，也就是状态