



人类和语言是共生的，是在互动中共同进化的。语言是思维的场所，是存在的家园（海德格尔语），其最基本的功能在于认知、交流和寄托。洪堡就认为一个民族的语言和思维是不可分割的，“一个民族的语言就是他们的精神，一个民族的精神就是他们的语言。”曾几何时，语言的作用被绝对化和泛化。萨丕尔-沃尔夫假说——所有高层次的思维都依赖于语言——更是把这种想法推向极致，认为语言形式和类别决定着人类的思维及产物。世界上各民族由其语言的不同，所以各民族对世界的认知和分析也不相同。不论是“语言绝对论”还是“语言相对论”，都有支持的案例，也有不一致的实验。但语言对认知的影响却是不可否认的。比如不同文化背景的人对于颜色有着不同的区分度，这主要受民族语言描述的颜色种类、文化环境所制约。举个极端的例子，据说爱斯基摩人可用两个明确的词表达雪山或雪墙在阳光下的“白”和在阴影中的“白”。还有，一个民族的语言在一种文化传统中积淀而形成的一个词汇实际承载着很多信息，在另一种语言中，很难被浓缩。比如，我们汉语中表示亲戚关系的“妯娌”、“连襟”等词，换成英文，须得若干个词方能表达清楚关系，但其中汉文化所传承的关于家族和亲情关系所赋予的信息我们就无法细论了。

语言不仅影响着个人的认知，同时也影响着不同人之间的交流，尤其随着互联网时代到来，语言已经

成为不同语种人之间交流的主要障碍。《圣经》有一个著名的巴别塔的故事，一开始工人们用同一种语言沟通，建造巴别塔的工程进行的非常顺利，但是由于上帝扰乱了语言、破坏了工程中人与人的交流，最终导致巴别塔建造的失败。这个故事折射出人们对语言障碍的深刻体验。

机器翻译、计算机语言处理在上个世纪就已开始，但那时，也仅仅是开始。虽然前辈先驱们有很多很好创意设想，但限于计算机技术以及人类对语言的认知阶段性水平，人类只是取得了有限的成就，只能做有限的语言处理。但今天，情况又大不一样了。互联网时代语言研究的一个重要特征是利用超机计算机、大数据、AI等技术来重新尝试解决语言问题，自然语言处理又升级到新的技术层次。科技情报分析处理的对象，如科技论文、科技报告、技术专利、科技新闻动态等，无不以语言文字为载体，如何充分利用自然语言处理技术，尤其是最新的深度神经网络技术，实现科技信息的分析、综合，完成信息向情报的转化，是实现情报研究和应用走向工程化的重要手段。虽然它还不能解决所有的问题，带来令人满意的结果，但阿尔法-Go已在围棋上胜过人类，那么计算机在语言和情报分析方面实现让人类初步满意的结果，也不会太久远。我们拭目以待。

本期《情报工程》共收录论文13篇，其中围绕机器翻译主题研究的论文9篇，主要介绍了近几年中日两国面向科技文献的机器翻译合作项目的相关经验与成果，内容包括神经网络机器翻译、基于实例的机器翻译、领域自适应、语料库建设等，另外还介绍了专利机器翻译系统融合方面的研究；另4篇则分别介绍了社交网络和旅游评价的情感分类、新闻文本的情感摘要和相似性检测等研究。随着深度神经网络技术的广泛应用，可以设想，语言处理及情报分析、数据修理技术正在开始新一轮的跳跃式发展，必然会带动情报工程实现更高层次的自动化和智能化，并将首先提高科技情报服务的水平。

刘靖超

2017年6月于北京