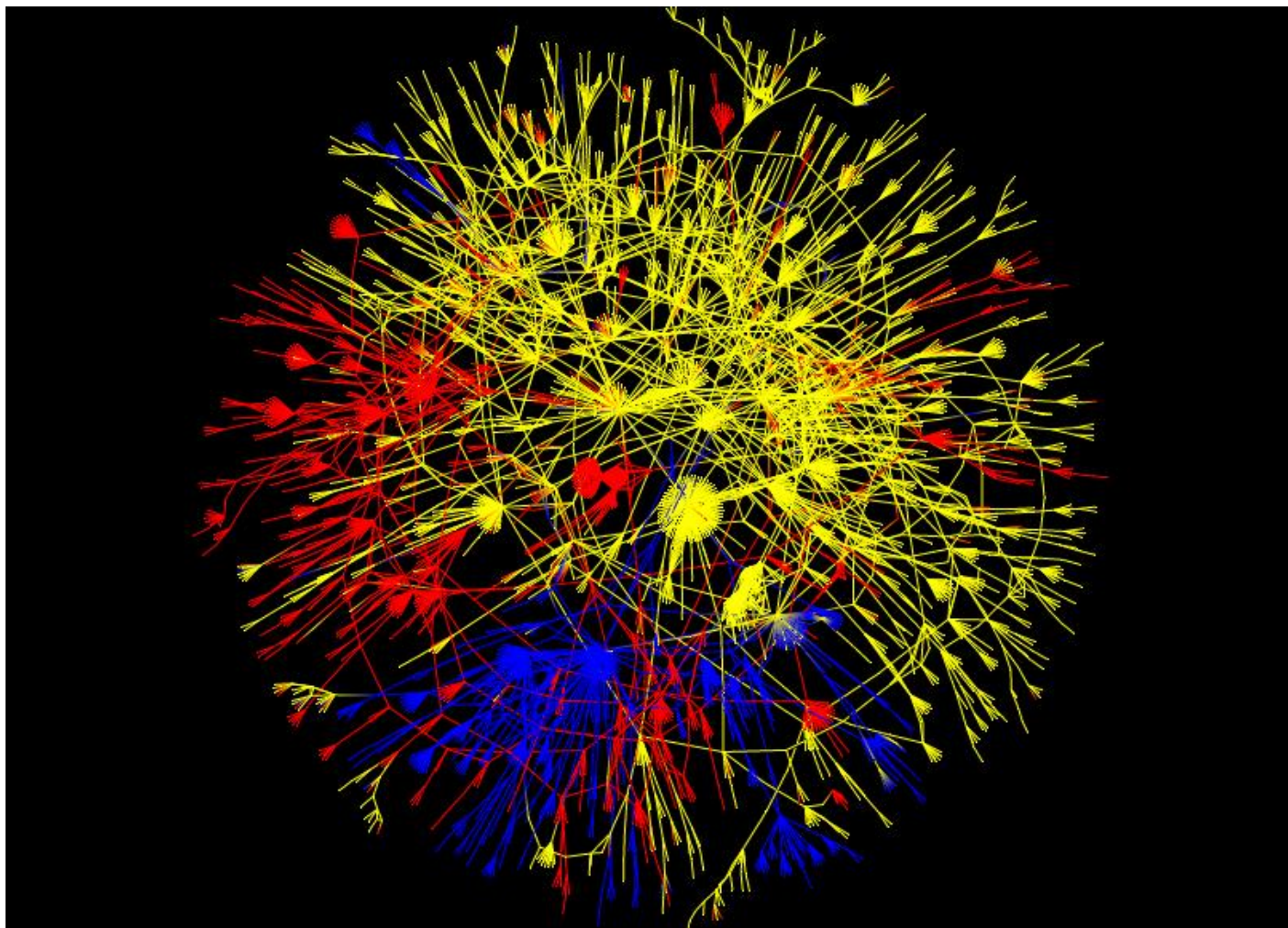


## 网路语言泡泡：最后的边疆？

Scott A Hale探讨了语言对在网上搜寻和传递讯息的影响。



### [言论自由大讨论的第一原则](#)

阐释了“无分国界，寻求、接受和传递各种讯息和思想”的权利。不过，语言是最明显却也最少人研究的国界。言论自由大讨论意识到了这一点，所以致力于把内容[翻译成十三种语言](#)这样浩大的工程。

语言到底会如何影响网络上的信息搜索和传播呢？现存的研究并未深究这个问题，而这也不是本文可以全面探讨的。不过，搜索引擎让我们可以一瞥不同语言间内容上的差异性。在搜索图片时，搜索引擎会试着把关键字和图片周边的文字、图片名和图片的链接文字配对。我们也许会认为，既然图片名往往独立于其描述文字，搜寻结果在不同语言间会大致相同。然而，图片的上传和注解取决于特定的社会语言背景。尽管谷歌不是所有市场中的主要搜索引擎提供者（雅虎日本、Yandex和百度分别在日本、俄罗斯和中国各自占有更多市场），但它在全球仍有领导地位，收录了大量的内容；在跨语言搜寻时，我们可自然推论，谷歌即使不使用完全相同的算法，也会用近似的算进行搜索容。因此，搜索同样地点，却在不同语言间得到不同结果，着实让人惊讶。

图一和二是用英文和中文谷歌图片搜索“天安门广场”的结果。所有的搜索都使用google.com，在英国的同一台电脑上先后执行搜索。然而，搜寻结果却大庭相径：这些差异有些并无大碍，但有的差异却让人担忧。例如搜索“天安门广场”的结果，显示出对1989年抗争事件的图片纪录，在中文和英文目录的数量有很大的落差。

图一：用英文在谷歌图片上搜索天安门广场的结果

图二：用中文在谷歌图片上搜索天安门广场的结果

其他[更有系统的研究，针对网路维基百科在不同语言版本间的不同，进行了调查。](#)

这些研究结果也显示“不同语言版本间重叠的内容少得惊人”(Hecht & Gergle, 2010)。最有趣的是，即使目前最庞大的版本，即英文版维基百科，平均来说也只包含了其他任何一版本讨论概念的60%（重叠率最高的是英文和希伯来文，比例达75%）。事实上，和第二大版本德文版维基百科相比，英文版只涵盖了德文版大约一半的概念；而德文版只涵盖了16%的英文文章。当然，即使“两种语言版本涵盖相同概念（非常清晰阐释），他们对概念的描述仍有可能有所不同”，[Hecht & Gergle \(2010\)](#)在他们的研究中对这个概念有更深入的分析。另外他们也提到如Omnipedia等新工具，可以让使用者探索不同语言版本间的差异。

这些现象并不都是负面的。事实上，有几个平台已经完全渗透全球（Facebook, Twitter, YouTube, 和Wikipedia只是其中几个例子）。这些平台上的社群沟通虽然离不了单一语言，但他们使资讯得以以前所未见的高速跨越国界传递。我的研究包含[2010海地地震后讨论此事的博客](#)（图三）和[2011日本海啸地震后，维基百科和Twitter上的连结分享](#)；再加上Irene Eleta对Twitter的研究，都显示了资讯散布横跨各语言的例子，也展现出有多种语言能力的使用者群有“桥结点”的功能，让资讯可以在不同语言团体间散布。

资讯在不同语言间扩散也有其科技和社会的面向

。 [机器翻译并非完美无误](#)

，但它仍成功使其他语言内容有高参与度；进一步的研究和训练资料也会持续增强机器翻译系统。

另外，有的研究—[包含我的研究](#)

—针对设计的角色，分析其如何影响使用者跨语言探索和传布资讯。这些研究都有其必要性。而社交媒体公司在建立新平台时，也应该利用这些研究并参考更广泛的社会科学研究的成果。当然，崭新的工具也会善用电脑和使用者的技术，在突破语言界限上有所收获。 [Doulingo](#) 和 [Monotrans2](#) 正是这种例子。他们让只会一种语言的使用者可以翻译内容。就Dounlingo来说，使用者还可以同时学习新语言。不过，[人类翻译](#)

的确有不可取代的地位；重要事件在不同语言中的讯息也要靠新的[媒体组织](#)辨其真假。

Scott A Hale是the Oxford Internet Institute的研究助理和博士候选人。

---

发表时间：七月 25, 2012