研究揭示中文不需要采用词间空格的认知根源

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/28178.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

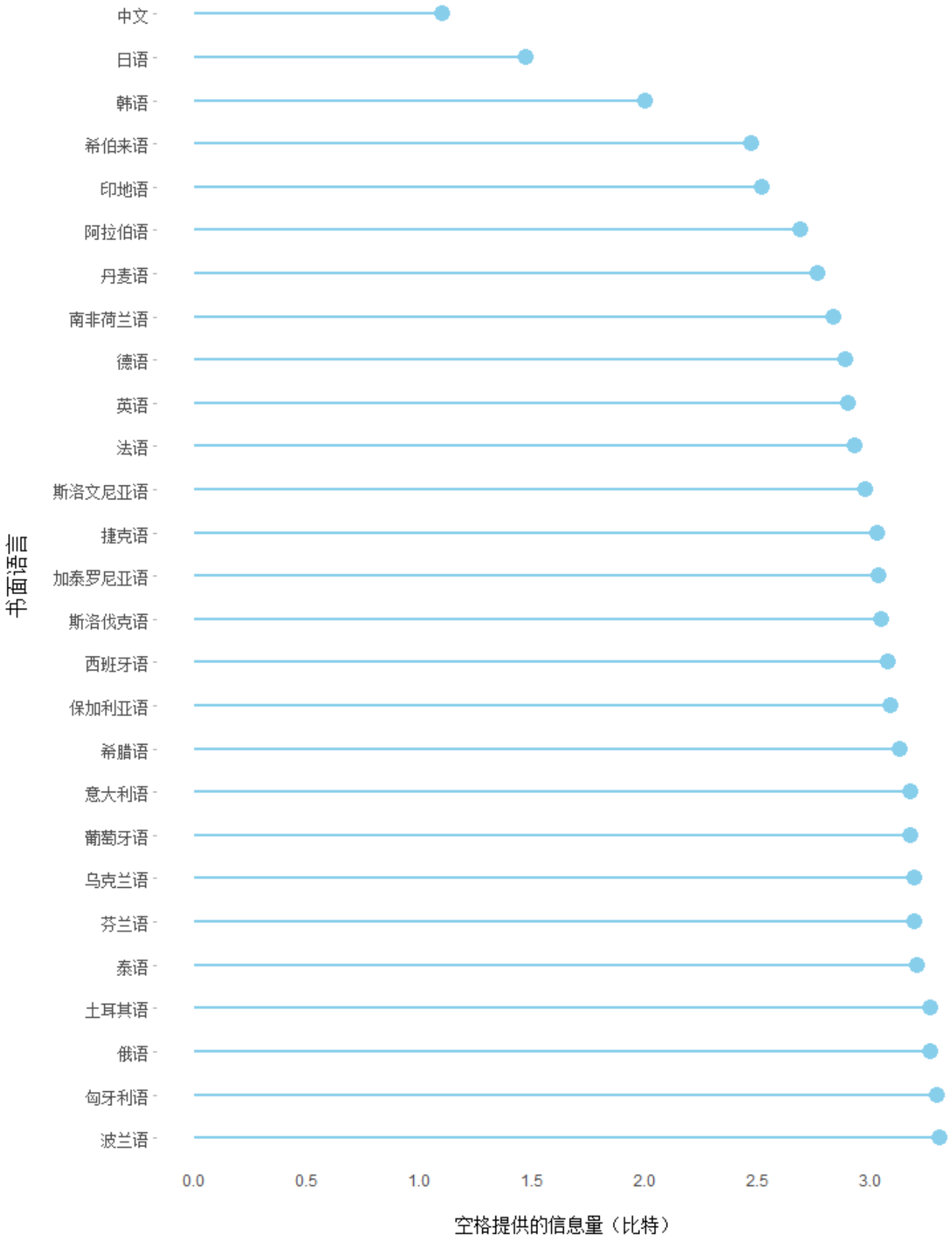
研究揭示中文不需要采用词间空格的认知根源

。英语中每个词之间都有空格，但中文没有空格将词隔开，属于不同词的汉字相邻呈现（如图1所示）。绝大部分的英文阅读模型认为利用空格将词分开在阅读中起到了至关重要的作用。这引发了一些重要的问题：为什么英语使用空格，而中文不使用？哪种方式更好？中文是否有必要在词之间加入空格，或者英语是否应该学习中文而取消空格？中国科学院心理研究所李兴珊研究组针对上述问题开展了一项研究。



图1不同书写系统标记词边界的方式

该研究提出了一个假设，认为书写系统对词边界标记方式的选择不是随意的，而是为了达到高效阅读，在权衡词切分需要付出的认知努力和阅读时的视觉加工效率后，选择了更经济的方式。为检验该假设，研究团队基于大规模语料库，运用信息论方法量化了27种语言中空格为确定词边界提供的信息量。结果表明，不同书写系统对词边界标记方式的选择与空格提供的词边界信息量有关：在采用空格的书写系统如英语中，空格提供的信息量更大（2.90比特）；而在不采用空格的书写系统如中文中，插入的空格提供的信息量更小（1.10比特，如图2所示）。空格提供的信息量反映了阅读无空格文本时付出的认知努力。对于空格信息量较大的字母书写系统如英语，文本去掉空格后，读者需要付出更多认知努力进行词切分，容易出现词切分错误；但对于中文，如果在文本中插入空格，空格提供的信息量较小，读者不需要付出太多认知努力进行词切分。因此，英语倾向于使用空格以减少词切分的认知负担，而中文则选择不使用空格。两者都选择了更经济的词边界标记方式。


书面语言
中文
日语
韩语
希伯来语
印地语
阿拉伯语
丹麦语
南非荷兰语
德语
英语
法语
斯洛文尼亚语
捷克语
加泰罗尼亚语
斯洛伐克语
西班牙语
保加利亚语
希腊语
意大利语
葡萄牙语
乌克兰语
芬兰语
泰语
土耳其语
俄语
匈牙利语
波兰语
0.0
0.5
1.0
1.5
2.0
2.5
3.0
空格提供的信息量（比特）

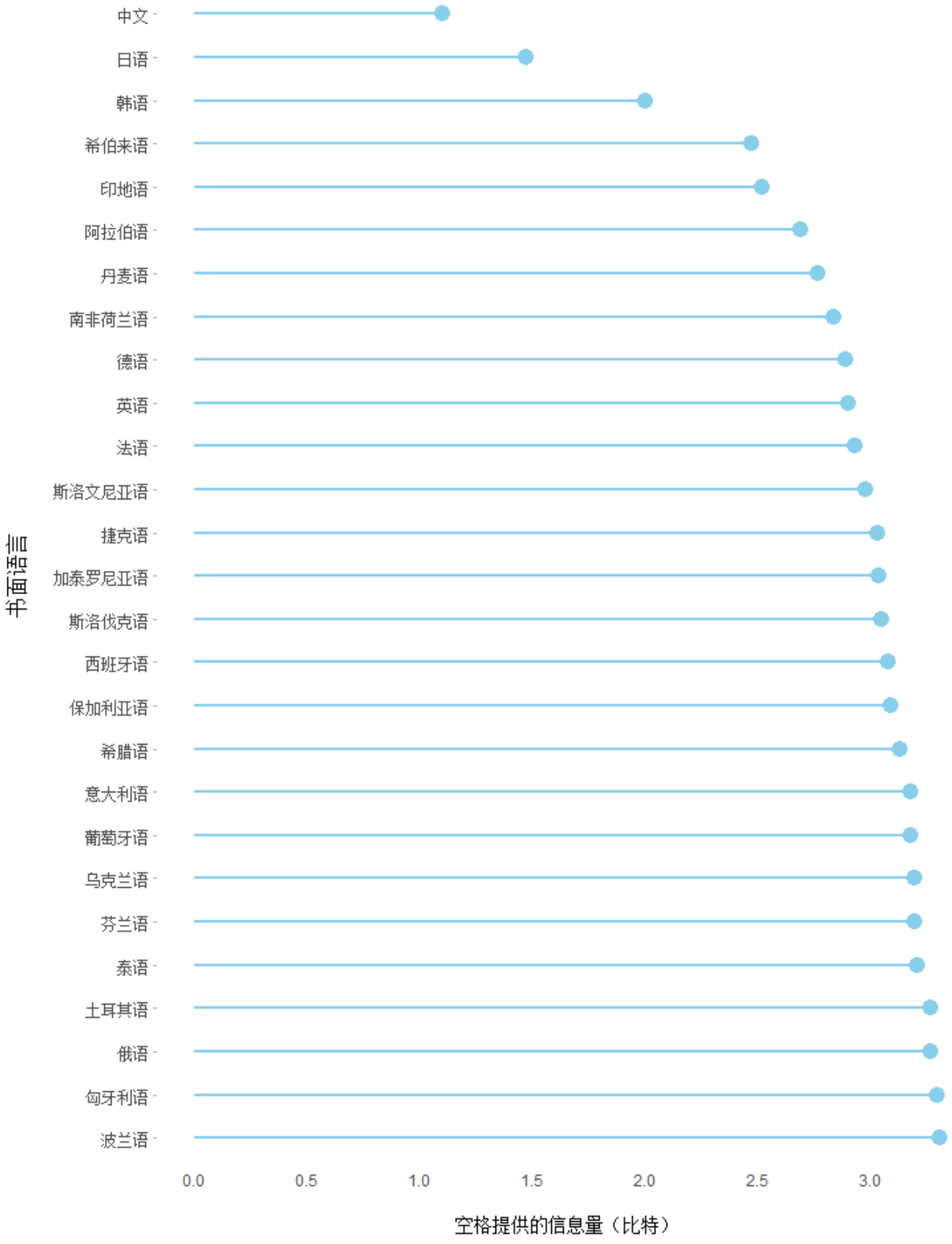

书面语言
中文
日语
韩语
希伯来语
印地语
阿拉伯语
丹麦语
南非荷兰语
德语
英语
法语
斯洛文尼亚语
捷克语
加泰罗尼亚语
斯洛伐克语
西班牙语
保加利亚语
希腊语
意大利语
葡萄牙语
乌克兰语
芬兰语
泰语
土耳其语
俄语
匈牙利语
波兰语
0.0
0.5
1.0
1.5
2.0
2.5
3.0
空格提供的信息量（比特）

图2二十七种语言中词间空格为确定词边界提供的信息量

空格信息量差异的根本原因是不同书写系统的词长分布不同。中文是表意文字系统，每个汉字代表一个音节或语素，因此每个汉字提供的信息量超过其他书写系统中的字母。中文汉字提供的平均信息量为9.84比特，而英语字母提供的平均信息量为4.32比特。基于该特点，大多数中文词可以用一到两个汉字表示（平均词长为1.40个汉字，标准差为0.57），而英语单词往往由多个字母组成（平均词长为3.78个字母，标准差为2.04）。因此，中文词边界位置的不确定性较小，即使插入词间空格，其为确定词边界提供的额外信息有限；而英语词边界位置不确定性较大，词间空格可以为确定词边界提供更多的信息量。

需要注意的是，中文若加入空格虽然提供的信息量较少，但仍提供了1.10比特的信息量。那中文为什么不用空格呢？这可能是在权衡了词切分付出的认知努力和视觉加工效率两个方面因素后，中文选择了更经济的词边界标记方式。在阅读时，一个注视点的视觉感知范围有限，空格的插入会导致读者在一个注视点上感知的字符变少，从而降低视觉感知效率。对中文而言，插入空格为词切分带来的效益不足以抵消它在视觉感知方面导致的代价，因此中文不采用空格是更经济的。相对地，英语等字母书写系统中空格提供的信息量较大，空格为词切分带来的效益远大于它在视觉感知方面导致的代价。由此可见，虽然有的书写系统采用空格标记词边界，而另外一些书写系统不明确标记词边界，但都是为了实现阅读的经济性而做出的选择。

与该研究的假设一致，以往研究表明，改变词边界的标记方式对不同语言读者的阅读效率产生了不同影响。这些研究发现，在空格信息量较大的书写系统（如英语）中去掉空格，阅读速率大幅下降约50%；而在空格信息量较小的书写系统（如中文），即使插入空格，阅读速率也不会显著提升（如图3所示）。

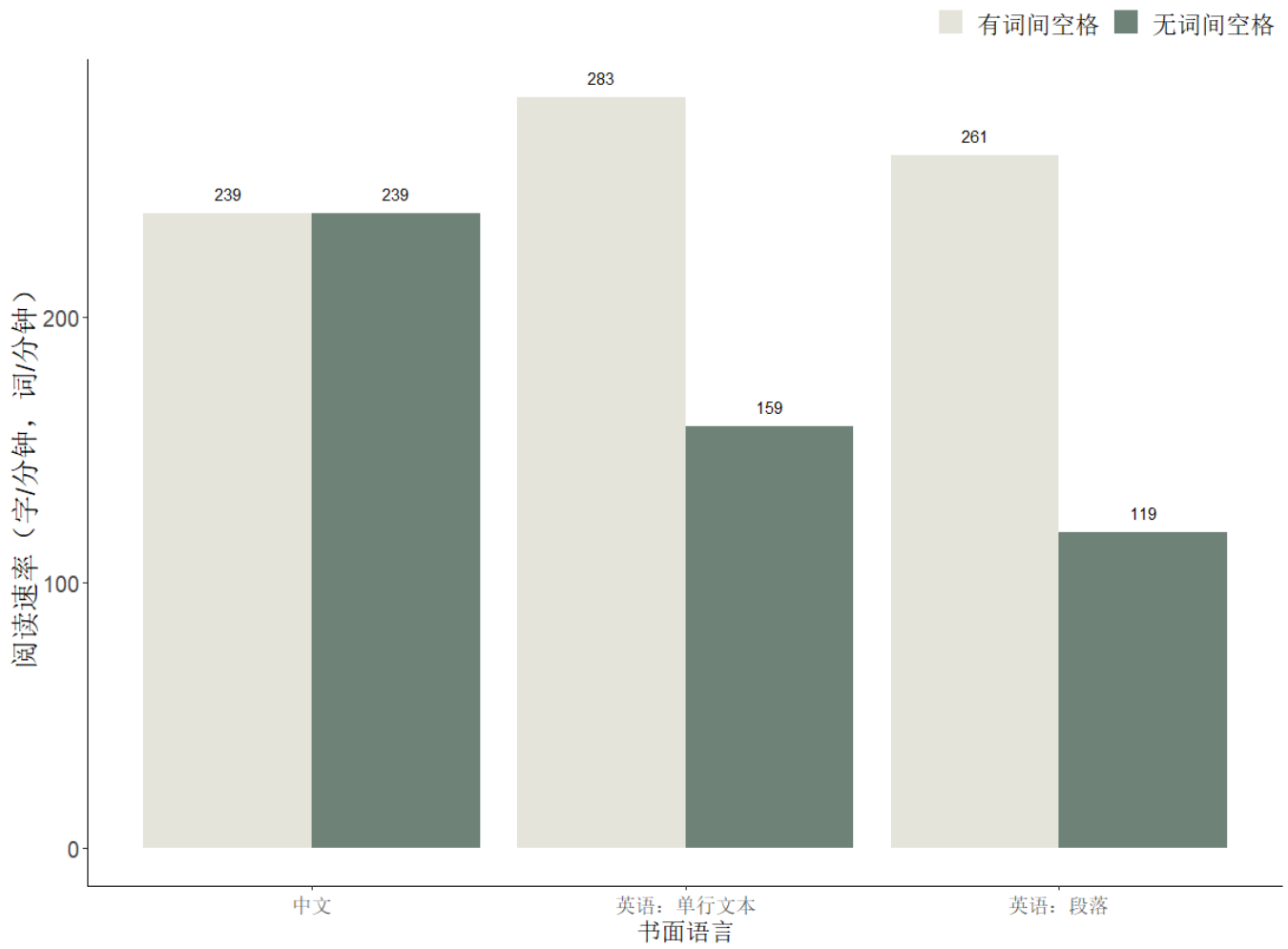


图3有词间空格与无词间空格的阅读速率

?

从字母书写系统的演化历史来看，人们逐渐改革书写系统，以实现最经济的词边界标记。历史上，字母书写系统并非始终使用空格标记词边界。由于转录的口语中没有词边界信息且书写材料昂贵，早期书写文本中没有空格。读者为了理解文本意义不得不出声阅读，导致阅读效率较低。这一时期的书写系统仅由少数抄写员或传教士使用。直到文艺复兴时期，随着大众阅读需求增加，这些书写系统才逐渐加入词间空格，提高了识字率和阅读效率。由此可见，字母语言中加入词间空格这一书写系统的变革逐渐适应了人类的认知需求，更符合经济性原则。

这一发现不仅有助于人们更好地理解跨语言普遍性与特异性的阅读认知机制，还可以为不同语言文本的排版和设计提供科学依据。

文章已在线发表于Annals of the New York Academy of Sciences。心理所助理研究员黄林洁琼为第一作者，心理所李兴珊研究员为通讯作者。该研究得到了国家自然科学基金面上项目（32371156）、中国博士后科学基金（2022M723362）、中国科学院心理研究所项目（E2CX6625CX）的支持

。

论文信息：

Huang, L., Reichle, E. D., Li, X. (2024/07/02).?Comparative Analyses of the Information Content of Letters, Characters, and Inter-Word Spaces Across Writing Systems.?Annals of the New York Academy of Sciences.?https://doi.org/10.1111/nyas.15178

相关论文：

Li, X., Huang, L., Yao, P., Hy?n?, J. (2022). Universal and specific reading mechanisms across different writing systems.?Nature Reviews Psychology,?1(3), 133-144.? https://doi.org/10.1038/s44159-022-00022-6

来源：中国科学院心理研究所

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发