

置顶 汉语是牺牲了什么，才成为世界最紧凑、最高效的语言？



黄风妖

本是后山人，偶作前堂客。醉舞经阁半卷书，坐井说天阔

1.4 万人赞同了该回答

汉语汉字是以“基础字符库大（要记的汉字多）、上手难度高”为代价，获取了高信息熵、底层架构好的优势。这两个缺点甚至直接导致汉字的两次拉丁化危机，使其差一点就像越南的喃字一样，泯灭于历史的长河中了……

一、汉字是世界最紧凑、最高效的语言之一吗？

答案是肯定的，以汉字为代表的意音方块字，作为一个复杂的文字符号系统，其信息熵很高。

从信息量的角度看，方块字与拉丁字母是二维码与条形码的区别。单从维度上说，方块字的信息含量就比字母文字高了一个数量级。

1948年，英国数学家克劳德·香农在论文《A Mathematical Theory of Communication》（通信的数学原理）中，提出了一个表征符号系统中单位符号平均信息量的指标——信息熵，并给出了计算信息熵的公式。

根据公式，符号系统中符号越多，信息熵越大；符号出现的频率越均匀，信息熵越高。据数理语言学中著名的齐普夫定律（ZIPF'SLAW）对全球主要语言进行统计，发现汉语是当今世界上信息量最大的文字符号系统。联合国五种工作语言文字的信息熵的比较如下：

- 法文：3.98比特
- 西班牙文：4.01比特
- 英文：4.03比特
- 俄文：4.25比特

汉字对拼音文字的这种信息熵优势是什么概念？

简单的比喻就是十进制数与二进制数的差别。十进制数字系统需要人记忆0-9，10个符号，二进制只需要记忆0和1两个符号。十进制乘除要记忆9X9表，二进制只需要学会与、或、非的简单逻辑。但是，人类在日常生活中为什么不使用二进制数字系统呢？因为那样很浪费，一个数字“7”表示成二进制就成了“111”，记个大数人不累？反过来，人类为什么不用十六进制，或更高的进制呢？一方面是人脑智力及生理习惯（人用10根手指来数数，习惯了）的限制，另一方面，十六进制也未必能大幅度提高信息熵。

汉字用5000个字解决了绝大多数人面对绝大多数文章的理解问题，就凭这一点汉字就已经领先拼音文字很多了。拼音文字要达到同样效果至少需要十万单词量，要一个人学10年以上。这种信息熵反映在文字上，就是联合国文件中，中文版本一定是最薄的。这就是汉字字符集信息熵高的优势。

而且如果觉得不够简练，汉语还贴心地提供多种压缩算法，例如：诗、词、曲、赋、成语、文言文……

这在古代书写材料稀缺的情况下（无论是西方的羊皮纸、莎草纸，还是中国的竹简、早期的纸，都是较为昂贵的），信息熵高的文字就更有利于知识的传播、保存，对文明的发展显得尤为关键了。

但世界万物，凡有所得，必有代价，正所谓“命运赠送的礼物，早已在暗中标好了价格”，那汉语汉字又牺牲了什么，才获得紧凑高效这一优点的呢？

汉语付出的代价就是““预加载字符库大（要记的汉字多）、上手难度高”，这两个缺点甚至直接导致汉字的两次拉丁化危机，使其差一点就像越南的喃字一样，泯灭于历史的长河中了……

二、汉语的缺陷与两次拉丁化危机

1 “学习上至难”导致的第一次拉丁化危机



1. 汉语的识字量（底层字符库）要求大，需要认识1500字以上才能脱盲。要想要能基本无障碍的阅读写作，差不多要能读写3500个左右的汉字，这个记忆量，对于初学者来说其实挺大的。而德文、韩文等拼音文字，本国的成年人学拼音文字只需花费几个月掌握数十个字母、音标、表音文字拼写和读音之间的规律后，便可快速入门，就可以做到会听就会念，会念就会写，做到“我手写我口”。
2. 造字法有点乱，汉字由于“言文分离”，其表音表意规律不清晰，很多汉字，哪怕是看起像形声字，由于几千年语音的变化，读音却与想象的相差甚远，识字不敢读半边，要靠死记硬背，这样一个个汉字背下来，可不得花两三年功夫。
3. 事实上，中文学习的难度大部分都在汉字上，如果不打算考虑读写的话，汉语其实是一门挺好学的语言，语法简洁，词汇的构成基本靠组合，反人类的发音不算多。对于拉丁语系的成年人，学汉字说是难如登天也不奇怪（如果只学拼音的话那就快多了，但哪怕只是学拼音，接下来的“四种声调”也够他们喝一壶了）……
4. 音调系统很古怪（这一条主要是针对外国初学者，这也他们最常见的抱怨之一，怎么可能Shuxue既可以是“数学”同时还能是“输血”呢？或者guojiang可以是“过奖”或者是“果酱”？）
5. 文化方面，汉语的难点在于，整个中文是依托于数千年的文化积淀而有序传承的，因此目前所用的词汇中有大量的典故、成语。（对于外国初学者）如果不了解其中的背景知识，那读起来是相当费劲的。

这是汉字的劣势，导致与其他文字相比，学习上手难度较高。因此，哪怕是本国的成年人要学习掌握中文常用的3500个汉字（占全部汉字使用率的99%以上），往往至少全日制学习三年才行，而拼音只需几个月就可以入门了（当然，仅限于本国母语，外语的话还是得苦学好几年的……）。

这也导致清末民初众多名人支持“汉字拼音化”的核心原因之一，当时这帮民国文化人的推理逻辑很简单：

中国积贫积弱->没有人才->受教育人少->汉字太难->拼音简单->汉字拼音化

因此，自“五四运动”之后的几十年里，无数文人志士致力于“汉字拼音化”，不是因为这些人吃饱了

3500个常用汉字至少要学全日制学三年，而拼音只需几个月。

而当时中国处于“亡国灭种”的正危急存亡时刻，什么都缺，连时间都缺。

亡国时刻的人们是没时间抒情的，没时间念旧的，更没时间学习茴字的四种写法。

他们迫不及待的想要在一夜之间培养四万万能读能写的人，这些人中可以诞生四千万能学习数理化的人，四百万学有所成的人，四十万高级工程师！建立四万所工厂，生产一个大帝国所需的所有铁甲舰和火车，化肥和石油！

而汉字拼音化，就是当时不少知识分子的想到的破局之道。

在他们看来，这不是愚昧无知，也不是崇洋媚外，而是生死关头的断臂求生。

能理解这些，就能理解为什么他们为什么要搞汉字拼音化了。

题外话：

多了解前因后果，会对历史人物的言论，多一份理解。但是，这份理解，绝不能等同于赞同。

虽能理解他们当时的所思所想，也能够体谅“他们要求拉丁化，全盘西化，本意还是为保存和发展种族”的本心，但若要给其一个盖棺定论的话，那就是——

这不过是一帮彻底被打碎文化自信、且不学无术（最多只学了半桶水）的“知识分子”的“病急乱投医”罢了。

二十世纪初那会儿，接二连三的战败早已打碎中国人天朝上国的自信，连“中学为体、西学为用”这种符合于当时文化自信的做法都深信不疑了。

在中国落后的现实下，中国的万事万物都有可能是造成落后的原因，所以病急乱投医，当真一厢情愿认为，只有否定延续了数千年的中华传统文化，否定了整个辉煌灿烂的中华文明，连汉字也放弃改成拉丁文，全盘西化“脱亚入欧”，才能救国救民。

说他们不学无术、半桶水，是因为他们根本没有做客观的科学研究，只是看别国强大，便随意臆测其强大的原因，譬如说拉丁化可以减少文盲，还不是一拍脑瓜的想法，真有什么科学依据么？

（扫盲的真正关键在于体系化的全民教育，而不是所谓拼音文字，你看亚非拉一堆用拼音文字的国家，还不是文盲遍地。反之如台湾，只要良好的全民教育体系，就算是用繁体字也能实现全岛扫盲。而一旦全民教育体系出问题，例如2020年代的美国，一样能整出全国识字率只有79%这种奇葩活）。

而真正脚踏实地的华夏先贤们，他们也知道为了救亡图存，就必须要提高广大群众的识字率。

所以他们在新中国成立后，面对识字率只有不到10%的现实，因地制宜制定的第一套简化字方案，大力倡导推广白话文，推行汉语拼音方案，在人民公社/生产队开展识字班，用最短的时间，最高的效率，让一个数亿人口的大国，消除了为数众多的文盲，适应了工业化大生产的需要，进而让一个沉沦落伍数百年的农业国，建成了一个规模齐全的现代工业体系。

这才是真的“救国救民，功德无量。”

2：“打字输入/印刷排版难”导致的第二次拉丁化危机

“打字输入/印刷排版难”这一点现在看来不算什么，但往前几十年，那真是一个极其严重的问题，汉字甚至差一点就因此彻底拉丁化了。

要知道，自清末以来大半个世纪的「汉字拉丁化⁺」思潮，其中一个重要理由就是——汉字印刷出版成本太高，阻碍知识文化快速传播普及。

而20世纪初，开启智慧与亚国家发展进步一个至关重要的环节，各国无不左和极普及教育，开展立

而汉语有5000多个单字，字库不仅比英文字库规模大，而且检索和排版也更为麻烦。任何一个出版社，小出版社要印东西最起码要准备好几百万个甚至上千万个字码，要不然就要临时刻字。而欧洲国家普遍使用字母文，字母大小写都算上，也不过几十个基础单元，日本的平假名片假名算一起，也100来个基础单元。相对而言，汉字印刷费时、费工、费料，印刷成本比拼音文字高指数量级。

之后印刷机和打字机的出现后更是雪上加霜，直接造成的结果就是——排版效率出现了巨大的差距，别人家出版业已经用了大半个世纪的机械化自动/半自动排版了，而直到上世纪80年代，我们还在人工排版。隔壁日本在19世纪就意识到了教育普及的重要性，于是为了适应西方铸排机的印刷方式，原本认为写汉字才高雅的大和民族立马向片假名滑跪了。我们呢？制版就靠工人在几千个铅字里一个个找，出份日报都费劲。你要是出本刊物出本书，里面生僻字再一多，猴年马月才能排完版。

出版业效率低，知识和信息就无法廉价高效的传播。

其实，当时的先贤们也为此努力过，挣扎过。因为汉字文字的打印和印刷问题，民国不少人呼号要改制中文打字机。1915年，留学生祁暄发明了一种中文打字机，使用了当时流行的“组合理论”。这台机器需要人工检索，效率极其低下。

1919年，商务印书馆工程师舒振东发明了“舒氏中文打字机”，该机器用了“常用字理论”。一般情况下用常用字模盘打字，有需要时再去生僻字里找。尽管这种方式大大降低门槛，但没经过培训也难以掌握。

1947年，民国的大师林语堂（曾在清华大学任教，著《吾国与吾民》向西方世界介绍中国），曾耗尽家财花费数年（从1919年一直搞到1947年）研制了第一台中文打字机——【明快打字机】，这打字机就和机械计算器的观感一样，当然也包含了机械计算器的劣势--超级贵;很精细，修理很难。因此这打字机根本没卖出几台，林语堂也因为这个机器破产了。

是的，由于汉字的特性，想要在机械时代弄出好用且低成本、易普及的打字机，是一件近乎不可能的事。

于是，当时的很多人（包括一些名人和文化界大人物）大多鼓励汉字拉丁化，因为这样我们就可以使用铸排机加快出版效率了。而且也有很大的群众基础在，毕竟大多数人本来就不识字，教他们学

1954年官方成立的文字改革委员会，委员会早就做了决定：“一定要实行拼音化”，虽然没有给拼音化下一个时间表，但是工作一直在做，就好比“一定要解放台湾”，没有时间表说明何时解放，但是一定要解放一样。

那什么时候中国官方才停止了推动拼音化的工作呢？

——是1985年，

为什么会在这时停止拼音化的探索呢？

因为1980年，咱们的王选院士发明了汉字激光照排系统，并利用该技术成功印刷了纪念周总理的文学作品——《伍豪之剑》。汉字出版业从此直接跨过机械化，实现了信息化，汉字输入/打印不再是问题了。

当时也是汉字最危险的时候，1980年代，是中国重新睁眼看世界的时代，而当时给国人的冲击，比清末更为剧烈，河殤文化被大部分人认可，不是没有原因的，所谓的反思，已经倾向否定一切，从文化到种族，从地理到血缘，中国的一切都被认为是荒唐腐朽，毫无意义的。

1980年代，一些小学教材已经改为全拼音化，整本书不见一个汉字，无数的老师虽不情愿但也只能私下里无奈地抱怨……

好在随着王选院士证明了汉字在信息时代一样能够快速输入/印刷，从而遏制这一浪潮。这些教材很快作废，学生们收到免费新的汉字教材，老师终于不用再抱怨了……



1980年王选院士发明了汉字激光照排系统

至此，王选院士用自己的努力，把【汉字拉丁化】彻底丢进了历史的垃圾桶。

(是那些有骨气的、坚信汉语的优越性、中国文明在计算机世界一定能找到自己的出路的人的坚持与创新才有了今天。有此国士，实乃中华之幸)

对于当时的人民来说，打字机时代如果一直延续下去的话，汉字就真的危险了——那种环境下，「汉字拼音/拉丁化」将会成为一种无奈的选择，被广泛推广……

但后来汉字激光照排和五笔输入法出现，汉字出版业直接跨过机械化，实现了信息化，解决了困扰近百年的难题，汉语拼音化随之被多数学者抛弃。汉字也就此躲过了一劫。

3、汉字的其他缺陷：

(1) 语素少，造成同音现象，听说层面信息传递低效

中文语素少，中文语素少，造成同音现象远多于英语、法语、德语等拼音文字。

同样具有位结构的英语时态变化、法语各种变位、阴阳性、因为单词变形成得清音比变化、在听说

也即，在听说层面，汉语的信息传递较为低效。这种低效导致汉语必须字正腔圆，逐字逐句，语速不能过快。而英语可以有闪音连读，因此在相同时间内能够传递更多的信息量。尤其在看视频时，没有字幕且说话人口音不标准时，就会不舒服。英语反而少很多。从这个角度看，汉语的简练是体现在读写上面的，而在听说层面显著落后于表音文字。这也很好理解，汉语作为意音文字，本身读写能力强大。英法语是所听即所见，记录的是读音（英语正字法太烂是另一回事），各有优势。简练也是看场景的。

同音造成的困扰，也是老外学习中文的一大苦恼，同音就罢了，同一个字，意思还那么多，简直是要让外国人分分钟崩溃。

前面说了那么多汉字的缺点，那么汉字真的不如英语等拼音/字母文字么？

当然不是!!!

汉字底层架构设计好，信息熵高（简练），哪怕是在全球文字中，汉字的底层架构是独一档的存在，他是一种拥有高度概念抽象能力的「面向对象」文字，拥有近乎完美的「封装」「继承」「多态」特性。每一个部首和基本汉字都是一个封装的「类(class)」，造字造词都是通过「继承」（最常用的是多重继承)来完成的，然后再通过声调完成「重载」以简化音系。

这些特性也让汉字成为了天生的跨学科文字，能以区区3500字便囊括表达万事万物能够有效解决了科学技术发展下单词量爆炸问题，让人最容易学到新事物和跨学科事物（大幅降低专业壁垒），让普通人也能快速高效读专业书籍。就这一点而言，汉字是可谓是“先天知识普惠圣体”啊，全民“通才”教育的至强神器，堪称最适合工业科技时代的文字之一……………

二 为何说汉字是“先天知识普惠圣体” 是适合工业科技时代的文字之一？



1. 可通过高效组词的方式，解决词汇量爆炸问题（掌握3500字，白万词汇都不怕）
2. 词组逻辑性好，容易理解（甚至“望文生义”也大差不差），大幅降低专业壁垒
3. 缩写词更易理解
4. “一字一音”的单音节模式，带来的更高效编码
5. 信息熵高（言简意赅，便于书写）
6. 具备“二维码”特性，便于认读（读得准+读得快）
7. 形音结合，解决同音字问题
8. 自带超大规模的“引用库”和压缩包

优势1：3500字走遍天下都不怕，有效解决词汇量爆炸问题，

所有语言都有一个特性，就是——你必须掌握它的基础量才能运用自如。比如汉语体系的字，英语体系的词。

那英文需要掌握多少词汇，才能满足基本生活需求呢？曾有专家专门整理的词汇频率表：

- 掌握2000英文单词，阅读覆盖率为75.89%
- 掌握9000英文单词，阅读覆盖率为98.10%
- 掌握24000英文单词，阅读覆盖率为99.65%，

也就是说，如果要满足98%的阅读覆盖率，就至少需要掌握9000个英文单词，要想满足99.6%的阅读覆盖率，就得掌握24000个单词，这个记忆的工作量着实不小啊……

咱拿中文对比一下：（数据来自国家标准 GB2312-80《信息交换用汉字编码字符集*基本集》相关词频统计）

- 掌握1000个汉字，阅读覆盖率为91.91%，
- 掌握2000个汉字，阅读覆盖率为98.38%，
- 掌握3500个汉字，阅读覆盖率为99.82%，

- 牛：bull
- 肉：meat
- 牛肉：beaf

[牛肉] 这个词，汉语就是 [牛] + [肉]，表达直接准确。而英语相对应的，应该是 [bull] +meat。但这样组合，就成了一个多重音节的词汇。对于表达和理解都十分费力。于是，英语就直接新造了一个词beaf，这完全就是资源的浪费。英语不仅需要记住 [牛] [肉]，还得记住 [牛肉]，而汉字系统只需要记住 [牛] 和 [肉] 这两个词汇就可以了。

而英文嘛，由于英文词汇之间无法建立像汉字词汇那样紧密的联系，就导致只要有一个新的概念在英语中出现，英文就必须设计出一种新词汇与之对应，而且很难在从字形看出彼此之间的联系，那要记的单词就多咯……

也就是说，对于英文零基础的中国人，要全职学习2年英文才能熟练掌握9000-1万个单词，雅思能考7-8分，这已经非常了不起啦，但去读报纸，小说，跨专业的学术文章时，会发现自己仍旧是个半文盲，还是满篇不认识的单词。

而中国人只要能掌握3500个汉字，以及相关词汇，就可以阅读《人民日报》而不会有多少生词。英语母语的人，得掌握24000个单词后，才能实现差不多的读报效果。

而且，英文读者一旦遇上报纸中那0.35%不认识的单词，那体验就不那么美妙了。前面说过，英文单词不像中文，难以“望文生义”，碰上了，不是认识就是不认识，猜都没法猜。按照0.35%的频率，平均每阅读1000个单词，就有近4个单词不认识。一页office文档就500个英文单词，看小说时每页都会出现1-2个不认识的单词，而若这是关键词，是非常影响阅读体验的。

所以，就最终效果而言，英国人学习了2万单词，他能享受的信息还是有限的。中国人学习三四千汉字，就可以享受几乎全部信息。可谓是“学少而知多”，投资收益率高的惊人呀。

同时，英文等拼音文字还存在一个问题：词汇量爆炸

举个极端点的例子，

目前世界上已知的植物有37万种。

拼音文字造37万单词也容易（字母排列组合下就行），，但是这也意味着，你要记几十万个单词。

谁能学会？谁能记住？这还不是最操蛋的。

更操蛋的是，世界上有150万种动物。

更更操蛋的是，世界上有数百万种微生物。

这还只是生物这一门学科……

如果每一个新事物都要造一个单词，那么造出来单词数量会吓死人，不可能有人认识。而且这些新词汇都需要学习者去重新记忆他们的拼写、音标、词义、时态变化，这样会造成英文词汇的学习记忆难度呈倍数级的上升！

而且这一问题随着时间推移，会显得越发突出。

- 1600年，莎士比亚时代（1564~1616年）的英语词汇量约为3万个单词（据《纽约时报》统计）；
- 1933年，第一版《牛津词典》编纂了76个年头，收录不到10万个单词；
- 1989年，第二版《牛津词典》，收录31万个单词；
- 2019年，据《纽约时报》统计，英语单词数量，超过了100万个；
- 2021年，趣词典网站统计，超过300万个。单词长度，平均在10–11个字母。**最长的正式单词1319个字母（其缩写词就好多了，缩写后仅仅28个字母）。**
- 目前，单词数仍在持续增长中（英语每年还有1~2万新词产生），看百万级的词汇量，估计普通人一辈子也记不完了。

而汉语是被冲击最小的一种语言。这核心在于汉语“四维结构”编码方式带来的高效组词优势。

通常来说，如果要对一类事物进行编码，大体上有两种方式，

- 一是设置较简单的编码体系（精简指令集），基础字符少，但组合成符码的长度会较长，
- 二是设置较复杂的符码体系（复杂指令集），基础字符多，但组合成符码的长度会较短。

英文采用的是精简指令集，使用26个字母（基础字符），形成“字母-词-句”的三维结构。

汉语采用的是复杂指令集，使用8种笔画构成3000~1万个字（基础字符），形成“笔画-字-词-句”的四维结构。

这种“四维结构”令汉语在组词方面，获得对英文“降维打击”的优势。

- 英文有26个字母（基础字符），两个字母可以形成676个组合，4个字母可以形成差不多45万种组合。
- 而汉语常用字有3500个，2~4个汉字组合的理论数量极大，2500个基本字可组成40万亿个词。

呃………怎么有种“26进制 VS 3500进制”的既视感？

汉语正是通过这样相对复杂的编码，构建了更多了的基础字符，从而可以通过更少字符的组合，为各种新事物生成表达准确含义的新词，避免了重新造字（神奇就神奇在你只需要了解最基本的汉字就可以懂得组合出的词的意思）。

客观世界是复杂的，而且随着科技的发展，人类对宇宙世界认知的不断提升，人类面对的各种概念会越来越多，越来越复杂。而这时候，英文等精简指令集的的不足就开始显现出来了，虽然你规则简单易上手，但人脑并不是图灵机，基础规则的便利（精简指令集）并不代表对人脑来说更方便。

举个简单的例子，



二进制的十进制呢？

数学上十进制的114514确实等于二进制的11011111101010010，但是在生理学上人的大脑最多同时处理5到9个信息(chunks)。原因是短期记忆储存空间的限制，超过9个信息团，将会使得大脑出现错误的概率大大提高。人眼一眼能识别的字符数也是有限的。

所以人们反而更喜欢日常使用学习成本更高的十进制。

同样地，三维语言由于维度较低，为了描述越来越复杂的客观世界，就必须创造越来越多的词。这给人们的学习生产生活带来了巨大的不便利，一个西方人要想较为熟练的掌握某项知识，需要掌握的各种词汇少则几万，多则十几万。一旦要转行学习新东西，还得重新背诵记忆。不仅难度门槛非常高，而且非常笨拙。

而汉语这种四维语言，除了可以用“构字法”来创造新字外，还有可以用“组词法”——由字的组合来产生词，由词来描述世界。

其实咱们的老祖宗当年造字，也曾走上过“构词（字）法”这条歪路，给每个事物都造一个新的对应的字/词（主要是当时还在竹简上刻字的年代，为了竹简轻点、字数少点，故意搞出来的一大把单字）。

比如马，少壮的马叫驹，高六尺的马叫驃，青黑色的马叫骐，毛色黄白相杂的马叫駉.....

后来老祖宗发现这样下去不是事呀。直接改用“组词法”来解决这一问题，比如将骐写成“青黑马”。从而避免这一问题。

这“组词法”的加持下，只要掌握了几千个汉字，这几千个汉字就可以组合成无穷无尽（近乎正无穷）的词汇，进而描述认知世界。非常简单而且非常巧妙。还可触类旁通，记忆量大减。

汉语里：

火箭，不就是火推动的箭么？

汉语这种强大的容纳、表述能力，可以利用不同组合就表达了不同事物，伟大而让人轻松。

英语呢？

有了火箭，就得产生个新词ROCKET；

有了计算机，就得来个COMPUTER。

记忆量大呀，挺累人的……

瞧瞧，这就是差距。

一边是要背诵掌握十几万个词，一边是要掌握几千个汉字，哪个简单，哪个先进一目了然。

优势2：高效组词便于理解，大幅降低专业知识门槛，可谓是“先天知识普惠圣体”

由于英文等拼音文字的词汇之间无法建立像汉字词汇那样紧密的联系，也很难在从字形看出彼此之间的联系，这就给普通人的阅读理解带来巨大的麻烦与不便。

举个简单的例子：

有个经典笑话：那就是【绝大部分英国人不知道葡萄干是用葡萄制作的，也不知道葡萄酒是用葡萄酿的】，为啥呢？

因为这几个货的单词是这样的：

- 葡萄 grape
- 葡萄干 raisin



再举个例子，为什么美国总统特朗普会给美国人民科普无人机没有人？

因为无人机英文是——drone，这五个字母组合在一起，实在是表达不出来——这东西里面没有人。

当然，也有读者会反驳——怎么会呢，这完全可以用词根词缀来解决啊。采取类似中文组词的方式，把词根作为构成单词的组件，利用词根来构成单词。这样即使你完全不认识这个新的单词，根据词根你可以猜出他的是个什么意思。

这看起来的确是好办法（毕竟中文就是这么干的），但是嘛……有些事别人（汉字）做起来简单，你（英文）做起来就不一定简单咯……

拼音文字如果要走这条路，代价就是——这样单词会变得很长。因为字母/拼音文字，由于缺少单音节词，所以一般常用词只能使用双音节词。而要形成逻辑结构就必须大规模使用四音节词，这种完全用词根词缀组合来表达海量的关联信息就太冗长了，且非常低效。

一个有意义的词根起码也得三四个字母，两三个词根的组合，这个单词就奔着十几二十个字母去了。如果稍微复杂一点的新词，学会这个单词的发音怕是得去德云社了。

举个典型的例子：

英文：pneumonoultramicroscopicsilicovolcanoconiosis

英 ['nju:mənəʊ ,ʌltrə-,maɪkrəs' kɒpɪk'sɪlɪkəvɒl'keɪnəʊ kəʊnɪ'əʊsɪs]

美 ['nju:mənəʊ ,ʌltrə-,maɪkrəs' kɒpɪk'sɪlɪkəvɒl'keɪnoʊ kəʊnɪ'əʊsɪs]

中文学名：肺尘埃沉着病（看着就像英文直译，有点拗口）

中文简称：肺尘病

所以现实中大家也基本不用这个词，而是使用silicosis来表述，但……短倒是短了，可它也就失去了望文生义的功能，非专业读者看到这个词恐怕想不到这是一种肺部疾病吧？

英文、德文里面的不少超级长的单词，就用上述这个方法造出来的，发音有规律，看了就能读。就是长了那么“亿点点”，没人会用。

而且随着时间的推移与不断的造词，导致英文单词越来越长，目前8个字母以上英文单词已经达到50%以上，甚至30%单词超过12位字母。英文单词平均长度，正在无休止的变长！

为了避免这种低效，很多常用的词就只能抛弃逻辑结构，只能造无法联想记忆的新词来避免单词的冗长与笨拙（又回到“一物一新单词”），然后嘛……就又跑回前面词汇量爆炸的老路了。

当然，还有一个办法，就是不造新词，直接借用现有词汇增加新解释，但这就会导致一词多义严重：

例如英文的set这个单词，大家可以看看它被塞了多少词义……

set - 百度翻译 >

☆ | 生词本

设置 常见释义

英 [set]  | 

美 [set]  | 

v. 设置;确定, 决定 (时间), 规定 (限度、指标、标准、界线、条件、价格、金额等);配置, 部署 (岗哨等);确定, 决定 (事件) 的发生时间;安装;使开始;使开始, 使着手 (做);分...

n. (互相关连的东西构成的) 一套, 一副, 一组;盘, 局;凝固;布景, 舞台; (尤指坚定的) 姿势, 姿态, 神情;沉落, 下沉; (因志趣、职业、地位等相同而相互交往的) 一群人, 一帮...

adj. 规定的; (书本等由教师或主试人为学生) 指定的;套 (餐);位于(或处于)...的; (天气) 持续的; (时间等) 预先确定的, 约定的;凝结的, 凝固的;有可能的;故意的;下定决心的, 坚决...

知乎 @黄风妖

- set, 有430个定义;
- go, 有368个定义;
- take, 有343个定义;
- stand, 有334个定义;
- get, 有289个定义;
- turn, 有288个定义;
- put, 有268个定义;
- fall, 有264个定义;
- strike, 有250个定义.....

不过, 平心而论的话, 这是其实是一个普遍问题, 只要是经过长时间的演化的文字都会存在这类“代码屎山”(毕竟任何项目经过数百年上千年的修修补补, 都会成为一坨屎山), 但得益于自身特性, 中文虽有一字多义, 【一词多义】情况却较为少见(尤其和英文对比)

从这点来看, 与英文这种拼音文字的屎山代码相比, 汉字的底层架构的先进性简直爆表了啊....

词汇量爆炸还带了一个严重问题: 专业词汇不断增多, 造成专业隔离。

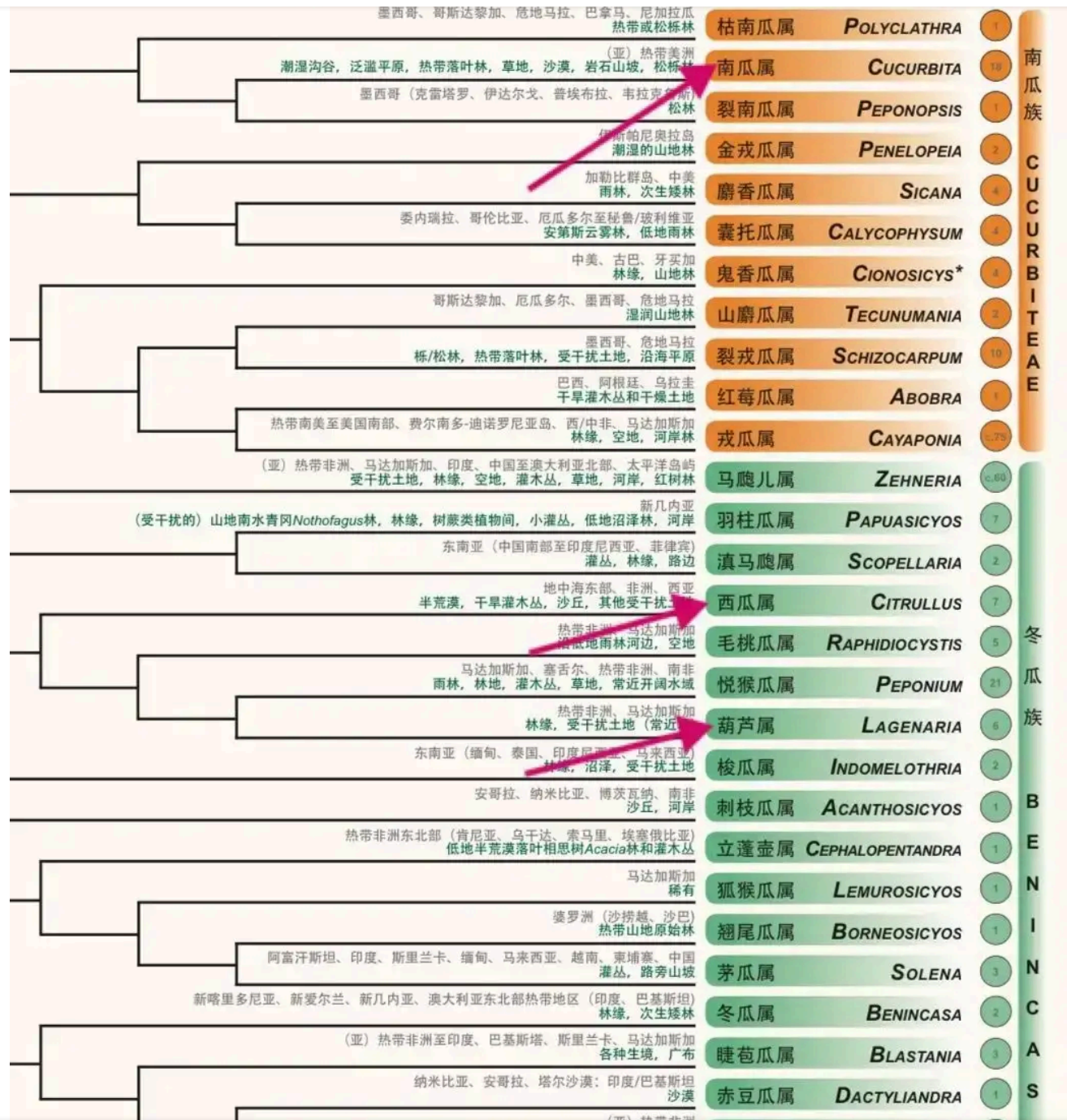
作为中国人: 你认识3500个汉字基本上可以读书看报了。

作为英国人: 你认识10000个英文单词还是个“半文盲”。

由于词汇量爆炸式增长, 在英文实际应用中, 每个专业都有大量专业英语词汇被发明出来, 并需要专门学习, 例如各种法律、医学、化学等专业都有数以千计的专有词汇, 为此还不得不专门编纂(各学科的)专业英语词典。

基本上, 任何一门学科都有专业英语, 而且其专业英文字典往往都非常厚, 如此一来, 造成的后果就是专业英语只有专业的人才可以记忆和使用, 导致英语的专业知识壁垒越来越高。而普通老百姓一旦遇到了哪怕稍微专业一点的学术技术领域名词, 立马懵逼变文盲了。

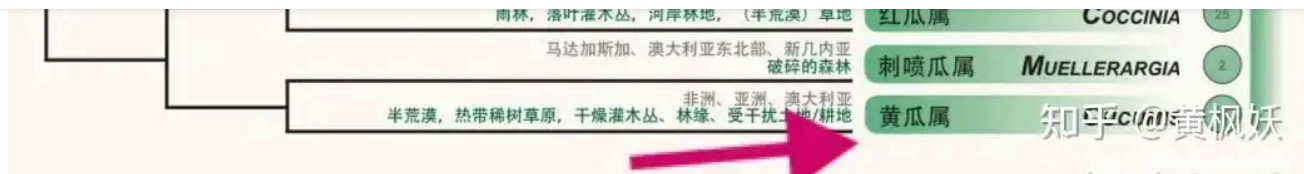




南瓜族
CUCURBITACEAE

冬瓜族

BENINCASINACEAE



看汉字部分，中国小学生也一看就懂，大概猜测这个瓜是啥特点。但是国外老百姓要是猛然撞见这些单词。。。立马听了懵B，见了流泪——完全不懂啊！

再比如医学领域，普通人学专业医学英语基本上是从头开始，跟以前学过的英语完全不一样。

不信的话，最方便验证的一个方法，去医院，你去看各科室的牌子，看英文标牌，你会觉得大学英语四六级白学了。

- 中国的医院里面有眼科，牙科，儿科。都是器官在加科字，就算是刚识字的小孩也能明白。但是英语就有点离谱了，child或者kid是儿童，但是pediatrics是儿科，牙齿是tooth但是dentistry是牙科，眼睛是eyes，但是眼科是ophthalmology。没个大学文凭，估计连医院科室都找不到。
- 在英美，如“四环素”、“变阻器”、“碳酸钙”、“高血压”、“肾结石”、“七边形”、“五面体”等词汇，都只有专业人士才会。
- 一般人连他们的基本医学术语，如“酒精绵球”“血压计”都不会讲。脑水肿、胃溃疡等病名，说给普通人听也听不懂，

普通外国人进了医院就只能听凭医生护士忽悠，连自己到底是啥病都不明白。

如果你对此还是难以体会的话，咱就用中文给你举个例子，你就明白“构词法”是如何制造所谓的专业知识壁垒，让你秒变“半文盲”的。

前面提到过，当年还在竹简上刻字的年代，咱们的老祖宗为了竹简轻点字数少点，故意搞出来不少单字。例如为了形容47种不同的马，造了47个字。

骚年，你知道“马”字的47种写法么？

5. 騃(xí):前脚全白的马
6. 𩇑(qián):四蹄全白的马。
7. 𩇑(céng):膝下白色的马,
8. 𩇑(yàn)屁股毛色白的马。
9. 騃(yàn):尾根白色的马,
10. 𩇑(láng):白尾马。
11. 𩇑(cōng)青白色的马。
12. 𩇑(tuó):有白色鳞状斑纹的青马+
13. 𩇑(zhuī):毛色苍白相杂的马
14. 𩇑(yīn):浅黑杂白的马,
15. 𩇑(bǎo):毛色黑白相杂的马。
16. 𩇑(quán):黑嘴白毛的马。
17. 𩇑(zhū):黑嘴的马。
18. 𩇑(zhān):脊背黑色的白马。
19. 𩇑(luò):尾和鬃毛黑色的白马。
20. 𩇑(máng):面、额为白色的黑马。
21. 𩇑(yù):股间白色的黑马
22. 𩇑(xí):背脊黄色的黑马。
23. 𩇑(diàn):黄色脊毛的黑马
24. 𩇑(guī):毛浅黑色的马,
25. 𩇑(xuān):青黑色的马。亦称“铁青马+”
26. 𩇑(róu):多鬃的青黑色马,
27. 𩇑(qì):有青黑色纹理如棋盘格子纹的马。
28. 𩇑(tiě):赤黑色的马。
29. 𩇑(lǐ):纯黑色的马。
30. 𩇑(xiá):毛色赤白相杂的马
31. 𩇑(yuán):赤毛白腹的马。
32. 𩇑(huá):赤色的骏马。
33. 𩇑(liú):黑鬃黑尾巴的红马,

38. 骠(biāo): 黄毛夹杂着白点子的马
39. 騮(guā): 黑嘴的黄马
40. 骃(yú): 紫色马
41. 馵(wén): 红鬃、白身、黄眼的马
42. 骄(jiāo): 六尺高的马
43. 駮(lái): 七尺高的马
44. 駮(róng): 八尺高的马
45. 驹(jū): 两岁以下的马
46. 駮(fēi): 三岁的马
47. 駮(táo): 三四岁的马

好家伙，我直接一个好家伙，马有这么多说法，如果要我说，我只会说一个泥码。

这要是搁古代。我妥妥就是文盲了。

所以这玩意在汉语造就被淘汰了。但这在英语中，可还是大行其道啊，尤其跨专业、跨学科阅读医学，生物，化学，法律的书藉文献时；众多的专业词汇，即使对于英语母语者来说，也是一辈子都没见过的生词，完全是在读天书一样。

这就导致英美等国出现一个严重问题——**功能性文盲**✦众多！

所谓功能性文盲，是指：“具有阅读、书写或计算能力，但是却缺乏利用这些能力来处理某些日常生活事务的能力”。“日常生活事务”指的是例如阅读报纸、查字典、读公车时刻表、理解说明书、填写申请书一类事情。

根据美国教育部和美国国家识字研究所2013年开展的【关于美国文盲人数的全面深入的调查】结果，美国有3200~5000万功能性文盲。

据调查结果，全美成年人中不怎么识字的占到了总人口的14%，也就是3200万人。全美21%的成人在人有阅读困难，达不到五年级小学生所应有的阅读水平，而美国高中毕业生中20%的人连自己

领域的)精英的大学生,照样会立马化身文盲。

而汉语新词是以现有的汉字组合起来,即使是科技时代很多新的专业词汇,汉字也可以通过组合的方式表达其复杂的含义。而中文由字组词,只用4千汉字可代表各种科学技术单词,解决了科学技术发展下单词量爆炸问题。

由于汉字的单音节特性,又有丰富含义,使其可以很奢侈地使用逻辑结构来构建“词组”。汉语新词是以现有的汉字组合起来,即使遇到没见过的新词,只要认识组成词语的单个汉字,你基本都可以望文生义,这种逻辑结构,使得中文的联系性、逻辑性要优于字母语言。

举例来说,

以英语为例,水仙花叫narcissus,如果该儿童不知道这个单词,他就无法根据这个词本身的组成部分推知它是什么意思。因为这个词的组成成分没有太多的常用的可以令人一望而知是什么东西的那种成分。narc有‘缉毒探员’的意思,narco有‘麻醉、昏睡’的意思。但它们一方面不属于最常用的词类,另一方面也不能暗示这个词本身是一种花。而且儿童也未必能把narc单独分离成一个词。

而在中国,当人民日报出现‘水仙花’这个单词时,中国儿童尽管也许不知道这究竟是什么花,但是他不需查字典就知道这是一种花。

水仙’两个字也是常用字,可以帮助+儿童判断这种花至少不是菊花、桃花、牡丹花及许多儿童已经知道的种种花卉。

也就是说,他即使不知道究竟是什么具体的花,他也已经知道

- 1) 这是一种花;
- 2) 这不是别的所有那些他已经知道的花。

知道什么是一种知识,知道什么东西不是什么,也是一种知识。如果将这个词中所含的信息量规定为3,则中国儿童已经不教而知其2,即67%可是,对于一个印欧语系的儿童来说,恐怕就不是这

下去，而印欧语系的儿童却会因为这个词而完全给卡住了，因为他如果压根儿就不知道这个词是什么意思，他就难以判断究竟它是一个具有关键意思的词呢，还是可以跳过不管的词。

这也是表音文字的缺点，就是出现新事物后，新发明的词汇意思和文字基本上没啥关联，你得不断扩展词汇量。而表意文字就是拿以前的文字凑合一下就是新词汇了，国内理工科专业课，谁见过需要专门去背词汇的？除非学的是英文教材。

实际上国内理工科本科学得认真点儿的，看大部分理工科研究生的论文，都能看懂个大概，跨专业交流，你说某人不知道其他专业的知识点是正常的，但要说完全听不懂人家说的啥，那倒是很罕见。

所以在中文互联网上，有时候双方互骂的时候会把对方比作小学生，甚至很多时候真的是小学生和大学生在网上辩论，虽然看起来这是个不好的行为，但反面说明了，中文的结构一旦理解，就非常容易突破功能性文盲的问题，即便是小学生，照样不存在能认字但无法用文字组成高级语句的问题。

因此，在英语国家，掌握1万单词的也就家庭主妇水平，真的什么工作也干不了，这才是对人最大的恶意，基本上你就会被锁死在底层。没有2万个词别想读报，没有3万个词别想把周刊读顺（据美国联邦教育委员会调查，美国中学毕业生七成看不懂纽约时报），大学毕业10年后的职业人士一般都得懂8万单词，才能从事知识密集型的工作，例如学个开挖掘机，先要学一大堆单词，学个驾驶，先要学一大堆单词，根本不利于科技普及。

而在现代社会一个说汉语的人只需要掌握3500个字（占汉字总使用率的99%以上），就可以读书、看报、搞科研。哪怕给你篇论文或来本专业书籍读，也很难找出不认识的字，即便是不认识的专业术语，相当一部分都可以望文生义。所以中国以后在多专业的交叉领域挺有优势的。

英语等字母拼音文字所谓的简单是入门比较简单，但仅能对付日常用语，却禁锢了使用者的知识摄入量，抬高了学习知识的门槛，是不利于全民科技推广的一种语言。

但英语也有一个好外 就是筛选（隔离？）能力特别强 中文的普及能力导致了 天才和普通人存在



中层面的英语词汇，绝对不足以支持你发表一篇比较有说服力的网络回答。

也就是说，即便是在中文环境中烂大街的辩题，如果你放到英语语境中，你只会收获很多非常夸张的阴谋论，而非一个看起来还算正常的网络辩论，能完成正常网络辩论的，一定是有一定学历水平以上的人士才能参加的工作。英语在功能性上的隔阂，特别容易把专业人士从普通人里面筛选出来，而英语本身的特性就是是个大筛选器，普通人是跨不过去的。

这也就是为什么英语国家下，民众和精英之间的对立程度特别大，比尔盖茨这样的人哪怕真的去非洲普及疫苗，也会被整上各种疫苗阴谋论，因为英语体系下，普通人和精英甚至可以说是两个物种，比如我们常说，（中文）每个字我都认识，但连起来我就看不懂了，在英文的表达环境里面是，你说的每个字我都不认识，连起来我更看不懂。

在疫情期间，为什么很多政策在外国无法执行，很多人喜欢说外国人崇尚自由，这是一方面，另一方面恐怕是，外国的普通人根本听不懂精英到底在说些什么。

同时，精英和民众隔离会导致他们的社会科学脱离底层，变成一种纯粹的精英游戏，很多时候你发现外国的各种民主制越搞越乱，其实本质也就在此，语言间的功能性隔离是跨不过去的，精英说的东西很多普通人一脸蒙，根本理解不了，比如上次脱欧，很多人都是等脱欧投票结束后才去查字典了解情况是什么，单纯的演讲根本传达不了足够的信息给普通人做判断，最终是精英获利，但责任甩给了普通人。

这点中文会有很大优势，民众能说很多只有外国精英才会说的词汇，而且能理解其意思，这就给上下信息流通带来很大便捷。

某程度上来说，汉字是可谓是“先天知识普惠圣体”，因为汉字天生就是属于跨学科的文字，让人最容易学到新事物和跨学科事物。让普通人也能快速高效读专业书籍，这才是最大的普世价值。

此外，汉语汉字这个特点也给中国工业/科技发展带来一个巨大优势——提供了一个超大规模的“通才”储备库

救急，“兵”能去读战术手册挖战壕打炮。

这就是为什么我对中国的世界工厂地位有信心。你在别的地方找不到这么大规模的“通才”储备库。你在别的语系，语种，语言环境下，找不到这么高效率的“基层从业人员培养体系”。这才是“世界工厂”的核心竞争力支撑啊……

优点3：中文缩写词更易理解

由于汉语具有远超其他语言的庞大单音节词库，因此在对特定名词进行缩写的时候，就更具有无可比拟的优势。表音语言在对名词缩写时，一般只能取首字母，这非常容易引起歧义。而中文可以直接取整个名词中比较具有代表意义的字，可以极大地消除歧义。

比如美国人说“国安局”就是“NSA”，只看这三个字母，不结合任何语境，大概很难猜到是什么意思；同样的，哪怕有语境，同样的缩写表示的意思可能会很多。但是中文缩写，基本上歧义率很低，对语境的要求弱，除非涉及高度专业性内容，不然读者一般都能看懂什么意思。你说个国安局，人即使不知道全称，国，安，局，三个字都已经涵盖了相关的意思。FBI这些不冷门的还能通过信息了解一下，那些冷门的怎么办，比如各种医学疾病的缩写，各种不出名的部门。在这些东西上就很容易感知知道中文有多么优秀。

换句话说，字母文字适合扫盲，但很难保持文化延续和科技交流。而汉语等意音方块字在历史文化延续传承和科技交流上简直不能再赞。

看大家回复挺多的，再来吐槽一下英文的缩写！

- 无人驾驶飞机，简称 无人机。英文Unmanned Aerial Vehicle，简称 UAV，百度百科有7个义项。
- 平视展示系统，简称 抬头显示，或抬显。英文Head Up Display，简称HUD，百度百科只有1个义项。
- 西北工业大学，简称西工大。麻省理工Massachusetts Institute of Technology，简称MIT，百度百科有8个义项。



看出来了吧?英文缩写的两个问题，一是含义完全丢失，二是重复率太高。这导致了一个问题，非专业人士，或者不联系语境，根本不知道缩写是什么。而汉语的简称通常都能保留下一部分含义，靠猜经常能猜出来，即使猜不出来，解释一遍也基本能记住了。

呵呵，大家看，中文相比英文的优势是不是就充分体现出来了？从这可以看出，类似长度的组合词，中文一般都要比英文的逻辑要更清晰。这不简简单单是效率的问题，而更是人民接受信息能力的问题。组词方式越短、越有逻辑性，学习就越简单。整个社会效率就越高。

汉字确实比字母难学一些，然而每个汉字的含金量非常高，每个人只要掌握三千常用字，就可以迅速地掌握庞大的词汇量，就能顺利阅读包括专业论文的各种文献（但是里面要是引用了大量英文可就看不懂了），这是中国教育文理科都能互相理解和挑剔、自学成才率高、平均智商高的一个重要因素。这也是我们的母语为每个人提供的学习便利。

优点4：单音节结构促进高效编码与传输

汉字是高度单音节化，发音趋于“一字一音”。这相对于拉丁文的多音节体系，汉语的单音节体系在传输效率方面具备很大的优势。

从信息论角度来考虑，编码是很有学问的。举个例子。我们知道计算机传输信息，实际上传输的都是0和1。那么，如果我们传输的各种信息出现的频率不一样高怎么办？答案是，出现越频繁的，编码越短。这样就能提高总体效率。

比方说，我们只有四种信息要传递。按一般的想法，自然是把这四种信息分别用00、01、10、11来表示。每个信息都需要用两位二进制数来表示，也就是说传播100条信息需要发送200个二进制数。但是如果其中有一种信息出现的概率是91%，而另外三种分别是3%。那么就可以使用另一种编码方式：1，01，001，000。平均下来这种传播方式传播100条信息需要发送 $91+2*3+3*3+3*3=115$ 个二进制数。显然比前面那种效率要高。

可以很奢侈地使用逻辑结构。这种逻辑结构，使得中文的联系性、逻辑性要优于一般语言。而与语音脱离的文字体系，则进一步支撑了这种结构，方便了记忆。

我们来举一个非常简单的例子。普通中国人的初等数学能力往往超过欧美。这并不简简单单是教育的问题。更关键地，这是中文对数字命名结果。

中文由于汉语在单音节词上无可匹敌优势，可以极度奢侈地给予每一个数字一个单音节发音。没有音调的语言，是不可能做到这一点的，因为还有其他更常见的东西需要占用宝贵的单音节词的资源。

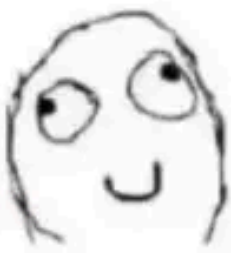
- 人类的喉咙发出的清晰且易分辨的单音数量是有限的，目前汉语发音包含音节（声母和韵母相拼）和读音（声、韵、调组合发音），查《新华字典》，普通话的23个声母(含y和w两个辅助韵母)和37个韵母(含er和ê两个特殊韵母)组合的音节有标准音节407个， 407×4 个声调=1628个读音，**普通话实际使用的读音（目前《新华字典》的单音数量）是1319个**，还有约300个读音没有汉字。
- 比汉语普通话单音数量多的，只有泰语、越南语了，泰语有5个声调，越南语有6个声调。（这也是因为汉语丧失了入声尖团音，在汉语单音在有入声尖团音的时代，其最大单音数约为2700）。
- **英语辅音28，元音20，由于没有声调，单音数量约400个，远不及汉语1300个多的单音数量。**

人对数字的短期记忆，实际上是对数字发音的记忆。研究表明中国人一次能够记住的数字长度要高于英语母语国家。而在计算中，你需要短期记忆很多数字，这一点就天然地给予了中国人绝对优势。

更进一步地说，中国的数字系统均基于单音节，这使得我们能够采用高度逻辑化的方式构建完整的数字体系。九十六，就是九个十加一个六。英语是“九十”（与九和十都不同的特殊词）加一个六。法语是四个二十加十六。汉语种最简洁而最富逻辑的结构，在世界各种主要语言中是独一无二的。

为了直观对比，可以看看下图，了解下各国语言是如何表达“97”这个数字的……

97



$90 + 7$



$7 + 90$

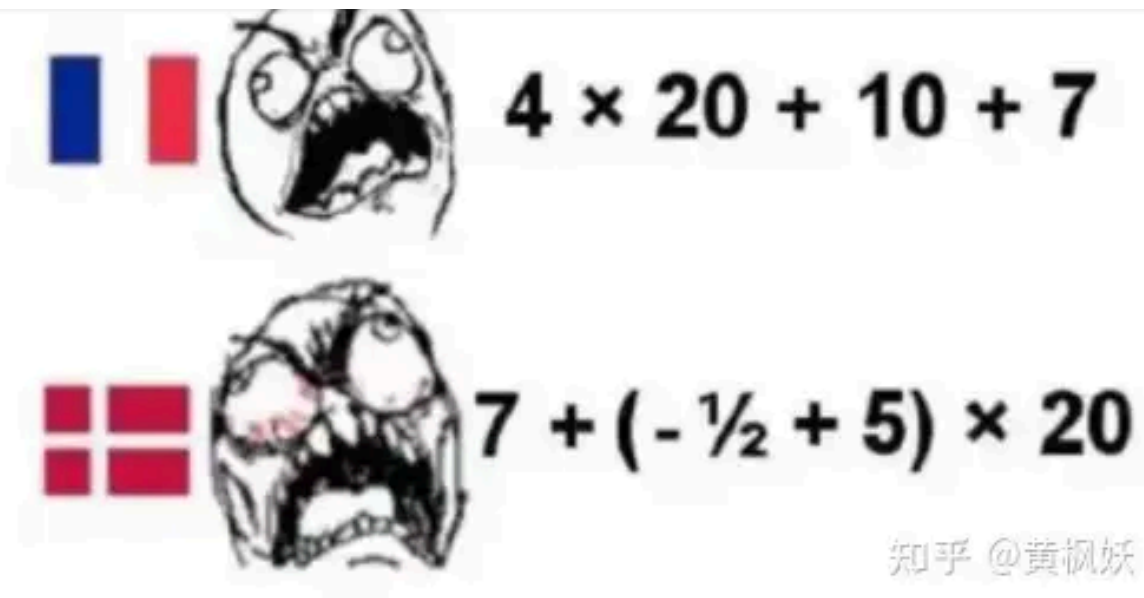


$9 \times 10 + 7$



$9 \times 10 + 7$





(吐槽：丹麦语 $97 = 7 + (-1/2 + 5) \times 20$ ，这个读法属实有毛病啊………)

九九乘法口诀表，就是构建在这个基础上的。其他国家的儿童如果想背下来这张表，可以说比中国儿童难了几倍。语言上具备了这种优势，中国人的初等数学怎么能不好？就初等数学上的优势，乘以初等数学在整个社会中的价值，这就是中国的根本竞争优势之一。

当然，需要指出的是，在较长单词的构成上，表音文字同样展现出一定的逻辑结构，这是语言发展过程中的必然现象。比如说英文nephritis，就来源于希腊语中一个读音类似的词，而希腊语中的那个词则来源于希腊语Nefros和itis，也就是“肾”和“炎”。不过很抱歉的是，由于表音能力差，这些很常见的意思必须要用这么多音节来表达。这就远不如中文简简单单的“肾炎”了。

优点5：信息熵高

前面已经说过，不再赘述

优点6：具备“二维码”特性，便于认读（读得准+读得快）

所以，本质上来说，拉丁系的表音文字（拼音）是一种一维线性符号（条形码），只有左右顺序有意义。

以汉字为代表的方块字，则是是二维码，除了左右，还有上下的排序。汉字能够在平面上沿竖直和水平两个方向延伸，所以一个汉字往往就能代表很复杂的意思，表达同样的信息，汉字也更省空间。汉字在更加充分地利用了人类的视觉系统，这使得读取速度大为加快。

此外，因为汉字充分的利用了平面的二维性，表达同样的信息时，在横向长度上比表音文字短。比如一个音节用一个汉字就够了，但是表音文字就要用几个符号把他的音素表示出来。而我们在阅读时同一时间视觉只能集中在一定的范围内，在这个范围内，我们可以看到汉字表达更多的意思。这也很大程度上提升汉字的阅读速度。

优点7：形音结合，解决同音字问题

汉字由于是意音文字，还具备一大优势，那就是避免了同音字太多导致无法识别的问题。

对付《施氏食狮史》、《季姬击鸡记⁺》这种同音文糊毫无压力（据说是当年某反对汉字拉丁化的学者撰写的小文，可自行百度）

《季姬击鸡记》

季姬寂，集鸡，鸡即棘鸡。棘鸡饥叽，季姬及箕稷济鸡。鸡既济，跻姬笈，季姬忌，急咭鸡，鸡急，继圾几，季姬急，即籍箕击鸡，箕疾击几伎，伎即齏，鸡叽集几基，季姬急极屣击鸡，鸡既殛，季姬激，即记《季姬击鸡记》。

《施氏食狮史》

石室诗士施氏，嗜狮，誓食十狮。施氏时时适市视狮。十时，适十狮市。是时，适施氏适市。施氏视十狮，十狮适十时，十时适十狮。是时，适施氏适市。施氏食狮史。

“放水”读音一致，结果建筑公司错误理解了图纸）。

优点8：自带超强的引用库

类比编程语言来说，拼音文字类似汇编语言，中文属于高级语言，并且自带几千年积累下来的各种函数和库。

汉语是依托于数千年的文化积淀而有序传承的，这令其有一个优势——有个巨大的引用库可调用，数以千计的成语、典故、诗词，为汉语提供着庞大的“超链接”。

调库不仅仅提供更高的信息熵，而且能让表达更加丰富有趣。

例如，一句“别在这亡羊补牢了”，显然讥讽力度要大于“事前有机会查缺补漏的时候你无动于衷，现在出了纰漏你在这做无用功”。这样成语典故在汉语中数不胜数，唾面自干、逐鹿中原、桃园结义、[千金买骨](#)、李广难封、冯唐易老、高山流水、知音难觅、桃李不言、三顾茅庐....上面这些成语基本都出自一两千年以前，但在今天的中文中仍然普遍使用。

当然，平心而论，任何一门语言，都会有自己的俗语、典故。例如英文、拉丁语中的潘多拉的盒子（Pandora's box）、达摩克利斯剑（the Swords of Damocles）、多米诺骨牌（Dominoes）等等。毕竟，任何文明经过了几百年了，谁家还没几个老祖宗传下来典故啊。只不过咱中国因为历史长了一点，所以咱的“引用库”也就比别家稍微丰富了那么“亿点点”……

不要小看这些浓缩的“超链接”资源，它在某种程度上塑造每一个中国人极其强大的战略战术及博弈策略素养，从耄耋老人到垂髫少儿，都是如此。

不信？

请阅读如下内容并回忆你什么时候学会这些词语的、并回忆自己是否运用过。

全蝤蛸壳 抛砖引玉 借刀杀人 以逸待劳 擒贼擒王 趁火打劫 关门捉贼 浑水摸鱼 打草惊



號。

兵者，诡道也。故能而示之不能，用而示之不用，近而示之远，远而示之近。

攻其无备，出其不意。

上兵伐谋 其次伐交 其次伐兵 其下攻城。

擒贼先擒王.....

短短几个字 能够清晰传递一个作战方案 因为每个词的背后都是千年传承历史经验。

几千年的纵横捭阖攻城拔寨 为中华文化留下了丰富的战略和战术经验 而这些又一步步的发展成为文化 潜移默化的留在了每个中国人的脑子里。

所以不论是受教育的军官还是不识字的士兵 都知道什么是以假乱真、出其不意、攻其不备、明修栈道、暗渡陈仓 然后穿插、绕袭、包饺子、张口袋.....

所以你看1962年对印反击战中，[庞国兴](#)这个三人临时结成战斗小组，就能一路“乘胜追击”、“声东击西”，将“风声鹤唳草木皆兵”的印军（整整一个营）彻底击溃，还打掉了敌人两个炮兵阵地。

菜市场里买菜大妈都知道欲擒故纵，嫌菜价贵，转身离开，而摊主也会来一招苦肉计，好好好我亏本给你.....

欧洲虽然也是经历过黑暗的中世纪，但其语言繁多、城邦较小，所以这样的通俗的谋略总结没那么多。而中华民族延绵不绝的5000多年的历史，浸润着每一个国人。

在这种环境下，每个中国人都着潜在的“智囊调用库”，碰到事情不知道怎么办，去翻一翻二十四史资治通鉴，保证有鲜活案例以及数套解决方案并且有成功经验失败教训。哪怕自己解决不了，找人商量也方便，“三个臭皮匠，顶个诸葛亮”嘛.....

汉字，自秦始皇“书同文”以来，本质上来说便成了东亚大地上的一种普世文字，深深地熔铸在华夏民族的血脉之中，其横则广布于万里之疆，纵则跨越悠悠千载，伴随着“华夏”这一普世帝国走过了一个又一个千年……

汉字是音形分离的，所有方言发音可以千奇百怪，但是写成文字一模一样，防止了形随音成在大地理区间和超长时间下形成新的语言，保持了语言的稳定不飘移分裂。

今天一个随便的中学生，就可以直接阅读，先秦诸子百家的著作，他可以自在地跨越千年，直接和历史上的先贤进行思想上的对话。

因为什么呢？

因为自秦以来，华夏便“书同文”了，每一个汉字都被赋予了一个相对稳定的意思。即便我们相隔千年，即便我们地域不同，口音各异，我们都可以自然而然地阅览秦文汉赋唐诗宋词元曲明清小说。

但是如果是一个英国人，他要读四百年前英文版圣经KJV 1611（King James Version，简称KJV 1611，英国国王请来诸多学者翻译的第一版英文圣经，最后成书于1611年）的话，就会发现，靠，这都写的什么呀，好古老的英文，表示看不懂，或者能看懂一点也懒得再往下看了。（KJV1611最难辨认的不是他的字体，就26个字母，字体你习惯就好了。最难辨认的是很多词的拼法和句子的用法，比较奇怪。），如果是更久远的古英语的话，那读起来就更牙疼了。

当然，这都还算好的，更为可悲的是文字传承彻底断绝，儿孙不识祖宗文，例如越南。

关于越南文字拉丁化的后果，大家可以看看知乎答主Yves S亲身经历的故事：

假如汉语完全拉丁化或是注音化后，会发生什么事情？

1.8 万赞同 · 1301 评论 回答