人工智能模拟辞例归纳的初步测试 *

莫伯峰　邱炜琦　谢泽澄

摘　要：古文字考释中的辞例归纳法，其实是综合了经验和理性两个方面共同作用的一种词义推定方法。人工智能语言模型现在主要模拟了人类经验主义的方法，并在日常语言处理方面取得了比较好的效果。如果将此类模型运用于古文字领域来模拟辞例归纳，也定会有所助益。我们基于 Bert 模型，用《四库全书》作为特定语料对模型进行了训练。以《上博简》(1-9)中 2103 个字为测试对象，模拟专家的部分辞例归纳能力，预测被遮蔽起来的文字。在总数 23157 的备选字符中，前 300 预测正确率达到 59%，前 100 预测正确率达到 46%，前 50 预测正确率达到 38%，前 10 预测正确率达到 25%，前 5 预测正确率达到 20%。可见，人工智能在古文字领域也具有类似人脑凭借语言经验进行辞例归纳的能力。同时，结果也提示，必须结合理性主义方法，才能实现完整的辞例归纳能力，建立相关的知识库必不可少。

关键词：古文字；辞例归纳；人工智能

辞例归纳是考释古文字的一种主要方法，前辈学者早有论述。唐兰在 1935 年写就的《古文字学导论》中称其为“推勘法”，将其列为考释古文字的第三种方法，认为“有许多文字是不认识的，但寻绎文义的结果，就可以认识了”①。黄德宽则更明确地提出，辞例归纳“是依据未识字出现的语言环境，通过对一系列辞例分析、比较、归纳，从而达到释字目的的方法”②。可以说，从事古文字考释的学者，有意识或无意识都一定会用到这种方法。

而实际上，辞例归纳不仅是古文字领域常用到的一种方法，只要使用语言，伴随着语言输出和语言理解，这种方法时刻都被运用着（只不过在其他领域，通常不把它单独提出来命名为“辞例归纳”）。这种“寻绎文义”“分析、比较、归纳”的过程在人类大脑里自主地进行着，显得如此自然而然，甚至很多时候我们都没有意识到它的存在。

只有当计算机程序复现“辞例归纳”的时候，我们才会注意到大脑的这一机能。比如计算机的联想输入法，只要输入一些内容，就能自动预测接下来可能会搭配的文字，这种预测其实就是根据之前的语境和辞例归纳出来的，也就是机器的“辞例归纳”。而如果使用语音输入法，输入的只是一组语音，输入法常常能根据这些语音拟测出我们想输入的文字，这种拟测也是依靠语音创设的语境和辞例而实现的。这一过程可以与人类的语言理解进行比较：聆听并记录他人的讲话，我们之所以可以把话语转写为文字，其实也就是在大脑中进行了“语音识别”的工作。所以，我们的大脑时时刻刻都在进行着辞例归纳。

现在，人工智能在处理日常语言方面，已经具有了非常强的辞例归纳能力。效果最为突出的是谷歌的 Bert，基于 Bert 的模型在机器阅读理解顶级水平测试 SQuAD1.1 中表现惊人，在两个衡量指标上均超越人类。在 SWAG 常识推理任务

* 本文系国家社科基金项目“利用神经网络进行甲骨卜辞字体分类的初步研究”（项目编号：19BYY171）的阶段性成果。

① 唐兰：《古文字学导论（增订本）》，济南：齐鲁书社，1981 年，第 170 页。

② 黄德宽：《古文字学》，上海：上海古籍出版社，2019 年，第 27 页。

中，Bert 模型也同样超越了人类水平[①]。那么将这种已经较为成熟的技术应用于古文字领域，来模拟古文字专家的辞例归纳，结果会怎样？前景会如何？这就是本文将要讨论的问题。

一、辞例归纳法是一种复合性方法

在开始让计算机模拟辞例归纳法之前，我们需要先论述一下，对于计算机而言，辞例归纳到底意味着什么？比较机器与人脑辞例归纳的过程，或许有助于更为深入地理解这一方法。对于人脑而言，通过辞例来限定词义的范围，看起来好像是一项很单纯的工作，仿佛由辞例我们本就应该能够卡定词义，不需要再做什么细致的讨论。但实际上，当用机器来模拟这一过程的时候，就会发现这一过程并不那么“单纯”。通过机器模拟我们可以发现，所谓的“辞例归纳”由两项能力联合组成，所以它是一种复合的方法。下面用一个简单的例子作以阐述：

《三字经》中的第一句“人之初，性本＊”。＊所掩藏的文字会是什么？相信绝大部分的人都会脱口而出＊是“善”。为什么“人之初，性本＊”的语境直接就能把我们带向“善”这个答案？如果再稍微思考一下，荀子还提出过“性恶论”，难道＊是“恶”的可能性就一定要比“善”小吗？由此可见，在根据辞例进行预测时，我们首先用到的是语感。在计算语言学中，将这称为经验主义的方法。

《三字经》中的第二句“性相近，习相＊”。＊号所掩藏的文字会是什么？假设我们从没读过这一句，没有任何语感可以借鉴。但相信大家稍稍琢磨，仍然可以猜出＊表示“远”，因为通过语境分析可以知道＊代表一个与“近”意义相反的词语，而根据我们脑中已有的常识和语言知识，便知道＊是“远”了。在计算语言学中，将这称为理性主义的方法。[②]

经验主义和理性主义的划分实际上反映了人脑思维的一般规律——“快思维”和“慢思维”两套系统各有千秋，各司其职[③]。大脑在处理各种信息时，常会将这两套系统综合地加以运用。当有大量经验可以依靠时，通常会用快思维来迅速处理。当没有太多经验可资利用，或者经验主义处理的结果不佳时，又会转向用理性的知识加以处理。

由此可见，在进行辞例归纳时，其实使用了两种截然不同的方法。经验主义会根据语感迅速缩小语义的范围，而理性主义则会进一步检查这些结果是否合乎理性。经过多轮次反复，最后确定答案。古文字考释中的辞例归纳，有的主要根据经验主义的方法得出结果，比如唐兰《古文字学导论》中论述“推勘法”提到的例子：金文中“眉寿”的考释主要就是通过经验主义的方法得出的[④]。时至今日，我们依然不能完全确定“眉”的含义，所以主要是依靠古代文献培养出的语感得出了这一结果。有的主要根据理性主义的方法得出结果，比如黄德宽《古文字学》中论述“辞例归纳法”提到的例子：甲骨文中的“𠂤”字，后世文献缺乏与之类似的辞例可以利用，难以完全凭借语感来推测，更多凭借了卜辞的上下语境，依靠常识和语言、历史知识进行推定[⑤]。

计算语言学在相当长一段时间里，曾将理性主义作为处理语言一种首选的方法，乔姆斯基的一系列语言学成就都与这种思想有关。这种思路在实际运用中效果总是不如人意，难以达到实用的级别，不过经过长期的研究却也产生了一套可以表示语言知识的方法，建立了一些语言知识库，比如 WordNet、MindNet、HowNet 等。而现在计算机进行自然语言处理，更多用到的是经验主义的方法，以概率模型而非确定性模型来处理语言，通过大量的语料让模型具有类似人类语感的

① Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv*:1810.04805(2018)

② 翁富良、王野翊：《计算语言学导论》，第一章第三节：计算语言学研究的基本方法（一、理性主义和经验主义：计算语言学研究方法的哲学分野），北京：中国社会科学出版社，2015 年。

③ [美国]丹尼尔·卡尼曼著，胡晓姣、李爱民 何梦莹译：《思考，快与慢》，北京：中信出版社，2012 年，第 5 页。

④ 唐兰：《古文字学导论（增订本）》，济南：齐鲁书社，1981 年，第 171 页。

⑤ 黄德宽：《古文字学》，第 29 页，上海：上海古籍出版社，2019 年，第 29 页。

能力，以词向量的方式来表示词语间的关系，这种方法在很多日常领域已经达到了实用级别的效果，只是在某些专业领域离人的水平还有一定距离。本次测试通过人工智能来模拟辞例归纳法，主要也是用了经验主义的方法，没有利用专家的知识加以辅助，所以是一个很初步的测试。

二、辞例归纳的理论依据、条件和效度

无论是人脑还是机器，为什么可以根据已知的辞例推导出未知的信息？这种机制与语言的性质有着密切的关联。索绪尔指出："语言是形式而不是实质。……语言是一种符号，它的意义产生于符号系统内部的关系中。"①这就告诉我们，决定语言中某个符号意义的，并不是符号本身，而是语言本身的系统。因此，辞例归纳模型背后其实蕴含了一套语言系统。过去理性主义曾希望由专家来搭建整套语言系统，目标是一套确定性模型，但效果有限。经验主义希望通过统计的方法，建立起统计语言模型，现在来看效果还比较理想。

要实现辞例归纳需要哪些条件呢？过去常提到的是"语境"。"语境"是一种显性的条件，所以很容易被察觉。比如一个"花"字，如果没有语境的限定，它到底表示"花朵"，还是表示"花费"，是无法确定的。语境越丰富，才可以越精确地归纳出词义。

但在古文字考释领域，大家所能利用的"语境"常常是相同的，就是有限的一些例句，却常常归纳出不一样的结果，这是什么原因导致的呢？这就涉及辞例归纳的另外两个条件——知识和语感，也就是前文所讲到的理性主义方法和经验主义方法各自的产物，它们都是隐性的条件，藏在我们的大脑里。

"知识"既包括领域知识，也包括常识。比如黄德宽论证辞例归纳法时曾提出："所谓语言环境，这里除指未识字所出现的上下文关系，还包括它铸刻的位置和使用的场合。"②"铸刻的位置和使用的场合"为什么能对词义卡定起到作用？背后就是有领域知识作支撑。而常识更是处处都在运用，比如林沄在论证辞例归纳法时所举的例子："（一个字）和雪字连用，就进一步又可以把红、绿等可能性也排除在外"③，之所以能把"红、绿"排除在外，就是根据一般性常识所得出的。现在，涉及"知识"方面的内容，人工智能主要是通过建立知识库来表达，本质是把人脑里的知识搬出来，用结构化的方式加以表示，以便计算机能加以利用。针对古文字及古汉语方面的知识库现在还非常少，所以这方面的工作还难以开展。如果希望计算机从数据中自动挖掘知识，也有一些初步的成果。但这涉及对语义的理解，理解是人工智能最大的难题，所有人工智能现在都还没达到"理解"这个层面，难以跨越"语义鸿沟"，因此效果还比较有限。

"语感"是一种通俗的说法，实质上是个体对语言这种符号的系统性掌握。黄德宽论述"辞例"时认为："（辞例即）词语按一定规则组成的序列，在这个序列中，各个词语是有机联系的，存在着相互依存和制约的关系。"④这里所说的"相互依存和制约的关系"表现出来就是"语感"。这种"依存和制约"关系是由语言的系统性所决定的，在一个系统里面，任何元素之间都会有依存和制约关系。这种"依存和制约"关系并不是确定的：一方面，语言系统本身总处在变动之中，所以这种"依存和制约"关系只能是概率性的。另一方面，个体对语言系统的掌握也存在差异，由此带来"语感"的不同，也就导致了辞例归纳出来的结果不尽相同。人类习得语感是一种很自然的过程，在一种语言环境中浸染久了，接触的语料足够多了，便自然具有了这种语言的语感。常常使用现代汉语，就具有了现代汉语的语感。充分接触古代汉语材料，才会具有古代汉语的语感⑤。本质上这即是一种概率统计的方法。现在，人工智能模拟人的语言能力，主要方向便是用概率统计的方法模拟这种语感。基于大量语料的训练，让计算

① ［瑞士］索绪尔著，高名凯译：《普通语言学教程》，北京：商务印书馆，2009 年，第 169 页。
② 黄德宽：《古文字学》，上海古籍出版社，2019 年，第 29 页。
③ 林沄：《古文字学简论》，北京：中华书局，2012 年，第 48—49 页。
④ 黄德宽：《古文字学》，上海：上海古籍出版社，2019 年，第 28 页。
⑤ 当然，由于现代汉语是由古代汉语发展而来的，所以掌握了现代汉语或多或少也会有一些古代汉语的语感。

机模型对语言符号的系统性产生某种认识，类似人类具有的语感。但是机器的这种语感跟人类的语感还有一些本质性差异，在本次测试中也有反映，下文我们将结合具体实例进行论述。

需要指出的是，辞例归纳并不能完全精确地卡定语义范围，也就是说它的效度是有限的，过去很多学者都已经充分地意识到这一点。比如唐兰就提出："虽然由这种方法认得的文字，不一定可信，但至少这种方法可以帮助我们找出认识的途径。"① 林沄也认为："辞例这一客观存在往往只是使我们在考虑不识的字为何字时，能有一定的范围，但并不是缩小到唯一的可能。"②所以，只是单独利用辞例归纳法，一定不能够得到明确的结果，无论是古文字学家还是人工智能，都是同样的道理。这也就决定本次测试的结果，必然也就只是划定一个大的意义范围，而不可能完全卡死。

由上述分析可知，语言的性质决定可以某种程度上模拟辞例归纳。但辞例归纳所需要的条件众多，所以机器现在还不能够实现完整的辞例归纳。而且，辞例归纳法的有限效度也决定它不可能独立完成文字考释。但是与人类相比，机器的信息处理能力无疑要强大很多，所以即使仅基于单方面的模拟，也可以期待在某些方面上达到人脑难以取得的效果。

三、模型的建立和测试过程

基于以上的认识，我们尝试用计算机建立了辞例归纳的语言模型，并对《上海博物馆藏战国楚竹书》(1—9)［以下简称《上博简》(1—9)］中的字词进行了遮蔽和预测，以模拟古文字学家辞例归纳的过程。以下对模型的建立和测试过程进行一个简要的介绍：

（一）模型建立的过程——从通用语感到古代汉语语感。我们选用了已经非常成熟的 Bert 作为基础，利用《四库全书》语料对其进行训练，实现让其具有古代汉语语感。具体训练过程如下：

首先，我们对《四库全书》数据进行了归一化。包括文本长度的设置，最大设置为 32 个字符（这是因为在预处理阶段统计《四库全书》文库时，发现大部分的句子长度都在这个范围内）。最终，处理得到的《四库全书》共计约 1900 万行文本，达数亿字符，统计汉字字符类别达 23157 个，外加五个特殊字符［PAD］、［UNK］、［CLS］、［SEP］和［MASK］。

其次，我们进行了模型具体设计。模型主要参数如下，编码层共计有四个隐藏层构成，隐藏层中 self-attention 注意力机制设置 attention head 数量为 4，中间特征数为 1024。同时，相关的 dropout 概率均设置为 0.1。考虑到我们的场景比较定制化，我们直接从头开始训练 Bert 模型。在训练过程中，我们采用的训练策略是，随机预测句子中 15%的单词。同时，在需要预测的单词中，将其中 80%替换成［mask］标注，表示这个字符被掩盖。10%的单词被随机替换成其它单词，而 10%的单词保持不变。

最后，基于 NVIDIA Tesla V100 显卡训练环境，我们耗时 45 个小时完成了 Bert 模型的训练，并进行快速部署和推理。

（二）模型测试的过程——以《上博简》(1—9)为测试对象。《上博简》(1—9)是出土文献中研究比较充分的一种材料，已经具有较为成熟的释文文本，材料性质也代表了古代汉语中比较典型的一种语料。因此，此次测试我们将其作为测试对象。

首先，我们对《上博简》(1—9)释文进行了预处理。由于辞例归纳的对象是词而非字，因此，释文语料是以词本位建立起来的。《上博简》(1—9)释文中的原形字被直接剔除掉，仅保留了文字所代表的词。

其次，我们对《上博简》(1—9)中字词进行预处理。包含同一文字的不同例句被提取出来（最多 11 例），并以［MASK］来替换这个将要被预测的文字。我们不以单一例句作为预测单位，是因为太过单一的语境难以取得良好的效果，如果说人类可以根据单一例句进行预测，那也更多是基于知识推导而实现的。

最后，我们让模型根据例句来预测被［MASK］遮挡的文字。预测范围设定为 300，超出 300 即表示预测失败。

① 唐兰：《古文字学导论（增订本）》，济南：齐鲁书社，1981 年，第 170 页。

② 林沄：《古文字学简论》，北京：中华书局，2012 年，第 48—49 页。

模型有效预测结果数据统计表

	前 300	前 100	前 50	前 10	前 5
数量	1242	974	810	518	428
比率①	59%	46%	38%	25%	20%

四、模型测试的初步结果及分析

模型建立好之后，我们对《上博简》(1—9)中的 2103 个字词进行了预测。在字符总数为 23157 的预测范围内，基本有效的预测结果见上表。

通过对各种结果进行的初步分析，可总结为以下五种类型：

(一)预测正确的词，特别是进入结果前 5 的词大多数是常用词，比如“于、为、一、万、上、下、不、世”等。以下选摘两则为示例：

1.“于”排第一位

＊＊＊ 于 ＊＊＊

0000 ： 及其见[MASK]外则物取之也

0001 ： 关雎以色喻[MASK]礼

0002 ： 以琴瑟之悦拟好色之愿以钟鼓之乐反纳[MASK]礼不亦能改乎

0003 ： 伐木实咎[MASK]己也

0004 ： 文王曰文王在上[MASK]昭[MASK]天

0005 ： 子曰为上可望而知也为下可述而识也则君不疑其臣臣不惑[MASK]君

0006 ： 诗云淑慎尔止不愆[MASK]仪

0007 ： 故言则虑其所终行则稽其所敝则民慎[MASK]言而谨[MASK]行

0008 ： 君奭云昔在上帝割申观文王德其集大命[MASK]厥身

0009 ： 昆龟筮犹弗知而况[MASK]人乎

0010 ： 关雎之改樛木之时汉广之智鹊巢之归甘棠之报绿衣之思燕燕之情盖曰动而皆贤[MASK]其初者也

"SUCCESS，于 at 0"

TOP 5 Prediction： 于之其也乎

2.“为”排第二位

＊＊＊ 为 ＊＊＊

0000 ： 民性固然甚贵其人必敬其位悦其人必好其所[MASK]恶其人者亦然

0001 ： 将大车之嚣也则以[MASK]不可如何也

0002 ： 敬宗庙之礼以[MASK]其本秉文之德以[MASK]其业肃雍显相

0003 ： 子曰[MASK]上可望而知也[MASK]下可述而识也则君不疑其臣臣不惑于君

0004 ： 子曰民以君 [MASK] 心君以民[MASK]体

0005 ： 诗云谁秉国成不自[MASK]正卒劳百姓

0006 ： 子曰上好仁则下之[MASK]仁也争先

0007 ： 此言之玷不可[MASK]

0008 ： 子曰宋人有言曰人而无恒不可[MASK]卜筮也

0009 ： 凡性[MASK]主物取之也

0010 ： 凡见者之谓物快于己者之谓悦物之势者之谓势有[MASK]也者之谓故

"SUCCESS，为 at 1"

TOP 5Prediction： 之为不以其

(二)一些不是特别常用的词，由于辞例较充分，效果也较好，比如“患、稽、俭、哀、嘘”等。以下选摘两则为示例：

3.“患”排第三位

＊＊＊ 患 ＊＊＊

0000 ： 青蝇知[MASK]而不知人

0001 ： 民之有惑[MASK]也上下之不和者其用心也将何如

0002 ： 凡忧[MASK]之事欲任乐事欲后

0003 ： 用智之疾者[MASK]为甚

0004 ： 无勉于民而治乱不[MASK]

0005 ： 举天下之名无有废者举天下之明王明君明士庸有求而不[MASK]

0006 ： 唯七年文王访于尚父曰我左[MASK]右难吾欲达中持道

① 为便于描述，此处比率数值取整数。

0007 ：民皆有决离之心而又有相[MASK]害之志是谓重殃重殃而贤者能以其六藏之守取亲焉是谓六亲之约

"SUCCESS，患 at 2"

TOP 5Prediction： 人右患之以

4.“稽”排第四位

＊＊＊ 稽 ＊＊＊

0000 ：故言则虑其所终行则[MASK]其所敝则民慎于言而谨于行

0001 ：耆老二拜[MASK]首曰朕兹不敏既得闻道恐弗能守

0002 ：夺民时以土功是谓[MASK]

0003 ：如欲察一仰而视之俯而揆之毋远求托于身[MASK]之

0004 ：守一以为天地[MASK]

0005 ：闻之曰一言而终不穷一言而有众一言而万民之利一言而为天地[MASK]

0006 ：如欲察一仰而视之俯而揆之毋远求托于身[MASK]之

0007 ：守一以为天地[MASK]

0008 ：闻之曰一言而终不穷一言而有众一言而万民之利一言而为天地[MASK]

"SUCCESS，稽 at 3"

TOP 5 Prediction： 之也顿稽为

（三）少量辞例不是很丰富的词，也预测出来了，但排名较靠后，比如“丕、严、仓”等。以下选摘两则为示例：

5.“丕”排第三十八位

＊＊＊ 丕 ＊＊＊

0000 :[MASK]显维德

SUCCESS，丕 at 38

TOP 50 Prediction：王不周文昭曰有以显天未弗光章明也先维○何武德无乃相莫思其而子威汤后大必之永康丕徐日烈尔三又方崇来福

6.“严”排第四十三位

＊＊＊ 严 ＊＊＊

0000 ：闻之曰从政有七几狱则兴威则民不道[MASK]则失众猛则无亲罚则民逃好刑则民作乱

0001 ：丘闻之臧文仲有言曰君子强则蹟威则民不道[MASK]则失众猛则无亲好刑则不祥好杀则作乱

0002 ：仰天事君[MASK]恪必信

SUCCESS，严 at 43

TOP 50 Prediction：行刑恪忠敬臣乱之子骄正不诚恶谨所恭宽顺仁也然是长出善死王至事慎淫俭过孝者大信理义诛心轻严公奢德道亡失

（四）预测结果不佳主要有两种情况：第一种情况是生僻字词，比如“悰、懋、剔、鬃、诎、轮”等。以下选摘两则为示例：

7.“悰”未排入前三百

＊＊＊ 悰 ＊＊＊

0000 ：有其为人之节节如也不有夫柬柬之心则[MASK]

FAIL，not in top 300

8.“懋”未排入前三百

＊＊＊ 懋 ＊＊＊

0000 ：初六旅琐琐此其所取[MASK]

FAIL，not in top 300

（五）第二种情况是辞例较少，比如“专、业、习、京”等。以下选摘两则为示例：

9.“专”未排入前三百

＊＊＊ 专 ＊＊＊

0000 ：王曰如四与五之间载之 [MASK]车以上乎

FAIL，not in top 300

10.“业”未排入前三百

＊＊＊ 业 ＊＊＊

0000 ：敬宗庙之礼以为其本秉文之德以为其[MASK]肃雍显相

0001 ：独居而乐有内[MASK]者也

0002 ：[MASK][MASK]天地纷纷而多彩物

FAIL，not in top 300

通过上述初步结果及分类来看，我们认为可以形成以下一些基本认识：

第一，古代汉语领域的人工智能辞例归纳会比现代汉语领域困难很多。

从初步测试结果来看，可以明显看到两种语料的效果差异。这是由二者的多方面差别决定的。首先，古代汉语语料背后所反映的语言系统性质并不单一。现代汉语是共时平面的同一语言系统，是大家都在使用的“活语言”。而模型所用到的《四库全书》语料是历时平面的语料集合，即使都属于文言系统，但背后所反映的语言系统也会因为时间差异而导致性质有所不同。更何况在这么长时间里形成的语

料，背后的知识系统也一定发生了很大的变化。这就决定了困难程度更大，如不加甄别地使用一定会影响效果。其次，古代汉语领域的字词关系等远比现代汉语领域复杂。古代汉语中的异体字、假借字、古今字等现象是现代汉语领域几乎不会遇到的，这些都带来了新的挑战。最后，古代汉语用字非常丰富，而且数据不均衡现象非常突出。我们所用《四库全书》包含了 23157 个不同字符，已经是一个较大的字库了，但是仍无法包含《上博简》(1—9)中的所有用字，而且古代汉语中所用到的文字远超 23157 这个数量，很多古代字典辞书的文字数量都远超于此。同时，不仅文字数量很多，而且这些字符的使用频率极不均衡，大量的生僻字词对于我们人类而言是学习的难点，对于人工智能而言也同样是一大难点。

第二，机器的“语感”与人类的“语感”存在明显的差异。

从以上的测试结果来看，模型所排列的预测结果会与人类专家的结果有明显差异。机器的各种预测结果之间缺乏联系显得比较凌乱，而人类专家的结果之间一定会有某种联系让这些结果具有可解释性。我们认为，二者的差异是它们对于语言系统的表示方式差别决定的，机器的语感是一种“纯粹”的语感，而人类的语感是经过理性提炼之后的混合语感。人工智能处理语言的基础是词向量，按照词向量的方式来理解词语间的关系，是将词语映射到高维空间，只能大致确定词语间的基本关系(见图 1)。而人脑除了上述这种模糊判定，还有很多明确的关系判定，而且具有清晰的层级关系(见图 2)。这种差异也就决定了机器的预测结果之间有时缺乏关联性，而人脑的预测结果之间一定是有某种关联的。

第三，辞例丰富的文字，模型预测会起到很好的辅助作用。

辞例丰富却难以敲定的字词，是古文字研究中的重点和难点。由于辞例丰富，常常可以对其词义有一定认识，但是要与后代具体的字词联系

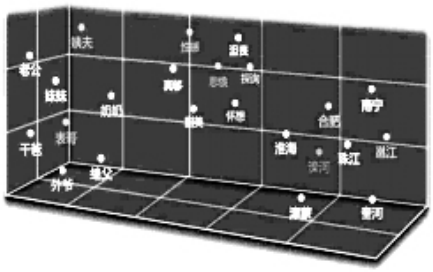

图 1　三维空间中词语关系的表示①

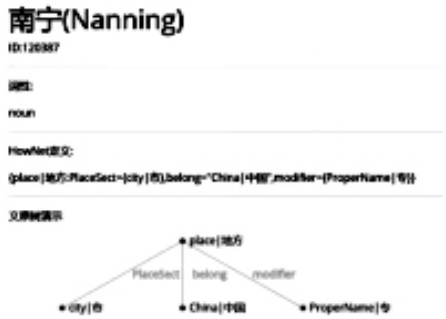

图 2　语言知识库中词语关系的表示②

上，需要有语感作支撑。这方面模型依靠强大的计算能力，具有一定的优势，如果能通过模型缩小备选词的范围将会大有裨益。

五、今后的发展方向

从以上的结果和认识出发，可以明显感觉到，只是基于语感无法全面地模拟辞例归纳，必须要有“知识”的介入才能让这项工作发挥出较好的效果。而这也是现阶段学界对于人工智能发展的一种共识。中国科学院院士张钹提出：“第三代 AI 其发展的思路是，把第一代的知识驱动和第二代的数据驱动结合起来，通过同时利用知识、数据、算法和算力等 4 个要素，构造更强大的 AI。”③长期从事计算语言学研究的冯志伟在很多场合反复强调：“应当把基于语义规则的理性主义方法和基于大数据的经验主义方法结合起来。”④语言学家陆俭明更是认为：“现在的状况是汉语本体研究成果没能在人工智能研究中派上

① 图片来源：https://ai.baidu.com/tech/nlp/word_embedding?castk=LTE=。

② 图片来源：https://openhownet.thunlp.org/entry?ID=120387。

③ 张钹、朱军、苏航：《迈向第三代人工智能》，《中国科学：信息科学》2020 年第 9 期。

④ 冯志伟：《神经网络、深度学习与自然语言处理》，《上海师范大学学报(哲学社会科学版)》，2021 年第 2 期。

用场。”①

如何将语言文字研究所获得的知识与人工智能相结合是一个很大的命题，从本测试的任务目标来看，如果把词语间的层次和意义关系加以确定，将会极大优化计算机辞例归纳的效果。比如“今”字：

11.“今”排第五位

＊＊＊ 今 ＊＊＊

0000 ： 子羔曰如舜在[MASK]之世则何若

0001 ： [MASK]纣为无道昏屠百姓桀约诸侯天将诛焉

0002 ： 六二王臣蹇蹇非[MASK]之故

0003 ： 仲弓曰[MASK]之君子

0004 ： 仲弓曰[MASK]之君子愎过攻责难以入谏

0005 ： 孔子曰[MASK]之君子宜

0006 ： [MASK]汝相夫子有臣万人道汝使老其家

0007 ： 不止曰小人之告将断于[MASK]日

0008 ： [MASK]君王或命毋现此则仆之罪也

0009 ： 太宰谓陵尹君入而语仆之言于君王君王之瘥从[MASK]日以瘥

0010 ： 君王之病将从[MASK]日以已

"SUCCESS， 今 at 4"

TOP 10 Prediction： 之人子曰今也我其是君

模型虽然在第五位就推测出了“今”字这一结果，但从前十结果来看，这种预测与我们人脑预测的结果存在比较大的差异。一方面，“之”是古汉语中最常出现的字，所以模型几乎所有的预测中都会排名靠前，而人类在预测时，如果发现被预测词辞例较少最先排除的也许就是“之”。本质上也就是用理性来对经验进行干预，这是模型还可以进一步优化的一方面。另一方面，人脑的结果一定会把同义词排在相近的预测位置，假如已经感觉“今”表示时间类的含义可能是正确的答案，那么我们一定会把“现昔向曩徂”等其他表示时间的词排列在基本相同的预测位置，这样就可以把一些生僻词的排序大大提前。当然，如果要让计算机实现这一目标，前提是要建立一套完整的古汉语字词知识库。只要有这样的知识库，相信模型的预测结果将极大地优化。

因此，从初步测试的结果来看，模型也许还难以直接对古文字考释发挥太大的作用。但是测试的结果却也给我们今后的发展指明了方向，即经验和知识的结合是人工智能模拟辞例归纳的必然要求。对于辞例归纳这项任务而言，如何将知识和经验进行结合，已经有了一些比较清晰的发展思路。

综上，通过初步测试，我们认为人工智能可以在某方面模拟辞例归纳，但要完整模拟辞例归纳，则必须建立相关的知识库。假以时日，人工智能模拟辞例归纳对古文字研究提供重大助力也是可以期待的。

（全文承厦门大学张俊松先生审读指正，在此深表感谢。）

作者简介：莫伯峰，首都师范大学甲骨文研究中心 ，出土文献与中国古代文明研究协同创新中心教授，近期主要从事计算甲骨学方面的研究。负责本文的项目设计和撰写。

邱炜琦，深圳中学国际体系学生。主要研究兴趣为语言哲学和中国古代文学，以及与人工智能技术的融合与应用。负责本文的数据整理。

谢泽澄，2019 年获得华南理工大学工学博士学位。主要研究方向为手写文字识别、自然语言处理和古籍文档分析。负责本文的模型搭建。

① 陆俭明：《亟需解决好中文信息处理和汉语本体研究的接口问题》，《当代修辞学》，2021 年第 1 期。后转载于“复旦新学术”微信公众号，题目改为《陆俭明：为何“人工智能对语言学的成果不是不需要，而是用不上”》。