

# 汉语预测语言学：理论与实践

## Linguistic Prediction in Chinese: Theoretical and Practical Aspects

孙敏 (Min SUN)

魁北克大学蒙特利尔分校教授（加拿大）

[sun.min@uqam.ca](mailto:sun.min@uqam.ca)

Mathieu GUIDERE

巴黎大学教授，法国国家健康与医学研究院研究主任（法国）

[mathieu.guidere@u-paris.fr](mailto:mathieu.guidere@u-paris.fr)

**摘要：** 本文将预测语言学的理论框架（Guidere, 2015, 2025）应用于现代汉语普通话的分析。汉语作为一种孤立语，具有严格的语序、高度的语境依赖性以及丰富的文化内涵，是预测分析的理想研究对象。在回顾源自大型语言模型的理论框架之后，本文构建了一套汉语预测语法，围绕三个层面展开：语义预测（基于词汇共现和语义偏旁预测下一个词）、句法预测（基于话题结构和体貌标记预测下一个短语或句子）以及语用预测（基于文化脚本预测叙事模式和会话进程）。这套预测语法为汉语教学提供了新的理论框架，同时为构建语言学上精确、文化上适切的预测性语言模型开辟了前景。

**关键词：** 预测语言学；现代汉语；预测语法；分布语义学；话题句法；文化语用学；语言模型；人工智能

**Abstract:** This article extends the foundations of predictive linguistics (Guidere, 2015, 2025) by applying them to Mandarin Chinese. It demonstrates that the Chinese language, characterized by isolating morphology, strict word order, strong context dependence, and a dense cultural fabric, constitutes a prime area for predictive analysis. After a review of the theoretical framework derived from major language models, the article develops a *predictive grammar of Chinese* structured around three levels: *semantic prediction* (anticipation of the next word based on lexical co-occurrences and semantic radicals), *syntactic prediction* (anticipation of the next phrase and sentence based on topical patterns and aspectual markers), and *pragmatic prediction* (anticipation of the narrative structure and conversational flow based on cultural scripts). This *predictive grammar* provides a theoretical framework for learning Chinese and opens up avenues for designing linguistically relevant and culturally appropriate models for the Chinese-speaking ecosystem.

**Keywords:** predictive linguistics, Mandarin Chinese, predictive grammar, distributional semantics, topic syntax, cultural pragmatics, language models, artificial intelligence.

## 引言

预测语言学由 Guidere (2015, 2025) 提出，是一个在当代神经网络架构背景下分析语言运作机制的新理论框架。其核心假说是：一切语言产出——无论口头还是书面——都源于基于先前语境的心理预期过程。在这一视角下，说话者和听话者持续调动一系列语言规律，使得某些语言接续比其他接续更具可预期性。因此，语法本身可以被重新定义为一套结构化这些预期的规则和约束系统。相关的语言单位不限于词：它们涵盖从音位特征到话语图式，中间经过语素、短语和句子等各个层面。由此，语言能力在本质上成为一种跨层面的预测能力，凌驾于传统的听说读写四项技能之上。

这一范式已在多种类型学上截然不同的语言（英语、阿拉伯语、日语、西班牙语）中得到了富有成效的应用，并揭示了各语言受形态学、语序和语用惯例制约的特定预测配置。将这一框架延伸至作为孤立语且使用表意文字的汉语普通话，开辟了尤为引人注目的研究前景。汉语不仅印证了预测语言学一般框架的适用性，而且转移了其理论支点，考验着其基本概念。

本文围绕三个方法论概念展开。

第一个概念是“预测单位”。在汉语这样缺乏显著屈折形态的语言中，预测单位既不对应语法语素，也不对应西方意义上的“词”，而是对应汉字（字，zì）。汉字的内部结构结合了语义偏旁（部首，bùshǒu）和声旁。因此，预测在一个字形、语义值和语音线索三者交织的层面上运作。

第二个概念是“语境窗口”，可以通过Transformer模型中的条件概率  $P(u_n | u_1 \dots u_{n-1})$  加以形式化。在汉语中，这一窗口不仅包括线性字符序列，还包括话题—述题结构、体貌和语气助词（了le、过guo、着zhe、吗ma、呢ne、吧ba）的存在，以及往往隐含的文化线索。

第三个概念是“预测层级”，它沿语言分析的主要层面展开：语义预测（预测下一个词汇单位）、句法预测（预测下一个短语或句子）以及语用预测（预测叙事图式或即将到来的会话进程）。这三个层级并非孤立的层次：它们相互制约，正是其相互关联构成了汉语的独特性。

除上述概念外，汉语还有三个结构特征使其成为预测语言学的极佳研究对象。

首先，分词本身就是一种预测行为：由于缺少正字法分隔符，词边界的识别成为一项概率任务，取决

于字符共现的频率。

其次，声调——四个具有区别意义的声调加一个轻声——执行着文字未标注但说话者的预测期望自然纳入的语音消歧功能。

最后，语境依赖度在汉语中达到了极高的水平，许多话语如果不调用共享的文化脚本——包括等级关系意识、面子（面子，*miànzi*）管理、礼貌仪式等——就无法被准确理解。

这种“预测性语境化”被构想为一种教学工具：它为汉语学习者提供了一种内化语言的新途径——不再是积累规则，而是逐步习得语境期望。

## 从一般框架到汉语的特殊性

预测语言学的基本假设是：形式和意义并非独立于其出现的语境而存在，而是共同涌现于由使用所稳定的分布规律之中。这一假说根植于二十世纪分布主义语言学的传统，如今在当代神经网络架构——特别是Transformer模型——中获得了数学形式化。预测被表达为条件概率 $P(u_n | u_1 \dots u_{n-1})$ ，其中每个语言单位的估计取决于由注意力机制加权的语境窗口。当这一框架应用于英语、阿拉伯语、日语或西班牙语时，揭示了各语言受形态学、成分拓扑结构和言语惯例制约的特定预测概貌。然而，普通话以独特的方

式对这一机制构成考验，甚至要求对预测语言学的概念进行部分重构。

第一个理论位移涉及分词。具有正字法分隔符的语言（如法语或英语）为自动处理提供了一个便捷的初级切分层，空格标示着词的边界。相比之下，汉语书面文本呈现为一串连续的字符流，没有明确的词界标记。无论是在字符层面、多音节词层面还是子词单位层面进行的分词，本身都成为一种预测操作：一串字符构成词汇单位的概率取决于其共现的频率和凝聚度。例如，字符串“中国人”可以被切分为“中国|人”（“中国的人”）或“中|国人”（“中部之国民”），只有语境及其激活的概率期望才能做出判断。因此，分词并非预测的前提——它本身就已经是一种预测行为。

第二个位移涉及语音学。汉语是一种声调语言，四个区别性声调加上一个轻声足以区分在其他方面同形或同音的单位。这一声调信息虽然不出现在日常书写中，但同样结构着说话者和听话者的期望。例如，最小对立组“妈”*mā*（“母亲”）、“麻”*má*（“麻”）、“马”*mǎ*（“马”）和“骂”*mà*（“骂”）充分说明了这一现象：消歧既通过声调预测进行，也通过语义语境进行。此外，大量非声调同音词使得语音预测与语义预测不可分割：学习者必

须学会“听”出文字所未写之义。

第三个位移更为深层，涉及语境性。诚然，任何语言都具有语境依赖性，但汉语将这一特性推向了极高的程度。主语省略频繁，回指隐含，句末助词承载着在其他语言中可能仅由语调传递的言外之力。尤其是，许多话语只有通过调用共享的文化脚本才能获得完整意义：等级关系意识、面子（面子，*miànzi*）管理、谦虚礼仪、礼貌和互惠仪式。一句简单的“吃了吗？”（*Chī le ma?*，字面意思“你吃了吗？”）不能被理解为关于饮食的真正提问：在大多数情境中，它作为一种寒暄公式发挥着类似于“你好”的功能。只有以对使用惯例的了解为基础的语用预测，才能消除歧义。

这三个特征——概率性分词、区别性声调和高度语境依赖性——并非汉语预测建模的障碍：恰恰相反，它们是最富生产力的机制。它们充分证明了为普通话构建专门的预测语法的必要性。

## 汉语预测语法

汉语预测语法与传统描写语法的区别在于其研究对象：它不试图阐述支配单位组合的封闭规则，而是旨在绘制使某些接续比其他接续更

为可期的概率规律。它不以“正确”与“错误”相对立；而是对“可能的”、“边缘的”和“出人意料的”进行梯度排列。在这一视角下，掌握普通话不再是记忆一套固定结构，而是逐步内化一个“期望视域”——该视域在三个互补层面上得以结构化：词汇单位层面（语义）、短语与句子层面（句法）以及话语与交际层面（语用）。

## 一、汉语的语义预测

语义预测关注在给定语境中最可能出现的词汇单位（字或多音节词）的预测。它依赖于三个交织的机制：词汇共现、偏旁之间的语义关联，以及语言中结晶化的搭配规律。

第一个机制——词汇共现——在构成现代汉语词汇主体的双音节词中尤为明显。当说话者说出“电”*diàn*（“电”）这个字时，可能的后续空间急剧收缩：“电脑”*diànnǎo*（“电脑”，字面义“电的大脑”）、“电话”*diànhuà*（“电话”，字面义“电的话语”）、“电影”*diànyǐng*（“电影”，字面义“电的影子”）、“电视”*diànshì*（“电视”，字面义“电的视觉”）。同样，在“飞”*fēi*（“飞”）之后，预期出现“飞机”*fēijī*（“飞机”）、“飞行”*fēixíng*（“飞行”）、“飞鸟”*fēiniǎo*（“飞鸟”）。语义预测在此通过具

有强分布亲和力的字符配对来运作。

第二个机制涉及表意结构本身。语义偏旁（部首，*bùshǒu*）构成一个强大的预测线索，在声旁被调用之前就已引导对字的解释。以水旁（氵）构成的字大概率属于液体现象领域：“河”*hé*（“河”）、“海”*hǎi*（“海”）、“湖”*hú*（“湖”）、“泪”*lèi*（“泪”）、“汗”*hàn*（“汗”）。以心旁（忄或心）构成的字指向情感或心理状态：“怕”*pà*（“怕”）、“忙”*máng*（“忙”）、“想”*xiǎng*（“想”）、“恨”*hèn*（“恨”）。这种规律性赋予了未知汉字仅凭偏旁就被正确解释的较高概率。

第三个机制是结晶化搭配，通常源于古典语言，其前半部分以近乎确定性的概率召唤后半部分。例如，“马马”*mǎmǎ*召唤“虎虎”*hǔhǔ*，组成“马马虎虎”*mǎmǎhūhū*（“马马虎虎”）。同样，“一石”*yī shí*在“一石二鸟”*yī shí èr niǎo*中召唤“二鸟”*èr niǎo*。成语（*chéngyǔ*）——四字惯用语——恰恰是汉语语义可预测性达到最高强度的形式之一：一旦前两个字产出，后两个字就受到强烈制约。

以下例句展示了语义预测的不同配置：

例1——语境语义场：在“我渴了，想喝一杯\_\_”（*Wǒ kě le, xiǎng hē yì bēi* \_\_）中，语境激活了一个

封闭的语义场：“水”*shuǐ*、“茶”*chá*、“咖啡”*kāfēi*、“果汁”*guǒzhī*。其他词汇上可能的接续在统计上极不可能。

例2——量词制约：在“一只”*yì zhī*之后，预期出现中小型动物：“猫”*māo*、“狗”*gǒu*、“鸟”*niǎo*。在“一张”*yì zhāng*之后，预期出现扁平物体：“纸”*zhǐ*、“票”*piào*、“照片”*zhàopiàn*、“桌子”*zhuōzi*。量词因此充当一种显性的预测过滤器。

例3——极性与词汇期望：在“他不但聪明，而且\_\_”（*Tā búdàn cōngmíng, érqiě* \_\_）中，后续被限定为褒义形容词：“勤奋”*qínfèn*、“善良”*shànliáng*、“努力”*nǔlì*。贬义形容词如“懒惰”*lǎnduō*将违反“不但.....而且.....”结构的连贯性和预测期望。

例4——对比预测：在“他虽然很累，但是还是\_\_”（*Tā suīrán hěn lèi, dànshì háishi* \_\_）中，让步结构“虽然.....但是.....”召唤表示努力或坚持的动词：“工作”*gōngzuò*、“学习”*xuéxí*、“坚持”*jiānchí*。

例5——成语：开头“画蛇”*huà shé*几乎只能接“添足”*tiān zú*，组成“画蛇添足”*huà shé tiān zú*（“多此一举”）。同样，“守株”*shǒu zhū*投射出“待兔”*dài tù*，组成“守株待兔”*shǒu zhū dài tù*（“不劳而获”）。

## 二、汉语的句法预测

句法预测关注延续正在进行的话语的短语或句子的预测。在汉语中，它依赖于一系列关联结构、体貌标记和话题图式，其预测价值很高，有时甚至超过屈折语言中所观察到的水平。缺乏动词变位在很大程度上被语序的严格性和助词的功能密度所补偿。

句法预测的第一个机制在于关联结构，其中第一个关联词的出现使第二个关联词的出现几乎成为必然。“因为……所以……”*yīnwèi... suǒyǐ...*是典型例子：“因为下雨了，所以我没去”*Yīnwèi xià yǔ le, suǒyǐ wǒ méi qù*。类似的还有“如果……就……”、“虽然……但是……”、“不是……而是……”、“只要……就……”、“越……越……”，它们在长距离上组织着句法期望。

第二个机制涉及话题—述题结构。如“这本书，我已经看过了”*Zhè běn shū, wǒ yǐjīng kàn guo le*，展示了话题前置，后接一个谓词结构受到高度制约的述题。一旦话题确立，系统便预期述题包含一个及物动词及其体貌修饰。

第三个机制是体貌助词和语气助词。动词后助词“了”*le*标志着完成，召唤结果补语或数量补语：“我吃了三碗饭”*Wǒ chī le sān wǎn fàn*。助词

“过”*guo*标志着过去的经历：“我去过北京”*Wǒ qù guo Běijīng*。助词“着”*zhe*标记状态的持续：“他笑着说”*Tā xiào zhe shuō*。每个助词选择一类可能的后续，并排除其他。

以下例句展示了不同的句法预测模式：

例1——条件关联：在“如果明天不下雨，\_\_”(*Rúguǒ míngtiān bù xià yǔ, \_\_*)中，结构投射出通常由“就”*jiù*引导的结论句：“我们就去公园”*wǒmen jiù qù gōngyuán*。省略“就”是有标记的，统计上罕见。

例2——递进关联：开头“他跑得越来越\_\_”*Tā pǎo de yuè lái yuè \_\_*选择一个可分级的形容词：“快”*kuài*、“慢”*màn*、“累”*lèi*。“越来越”结构排除不可分级的形容词。

例3——把字句：在“他把门\_\_”*Tā bǎ mén \_\_*中，该结构将确定宾语前置化，并投射出一个结果性动作动词：“关了”*guān le*、“打开了”*dǎ kāi le*、“修好了”*xiū hǎo le*。状态动词（如“喜欢”*xiǐ huan*）被排除。

例4——被字句：话语“他被老板\_\_”*Tā bèi lǎo bǎn \_\_*召唤具有强施事极性的及物动词：“批评了”*pī píng le*、“表扬了”*biǎo yáng le*、“辞退了”*cí tuì le*。

例5——话题前置：话语“中国菜

，我最喜欢的是\_\_”Zhōngguó cài, wǒ zuì xǐhuan de shì \_\_ 召唤一个菜名：“北京烤鸭”Běijīng kǎoyā、“麻婆豆腐”mápó dòufu、“饺子”jiǎozi。

例6——正反问句：A-不-A结构，如“你去不去？”Nǐ qù bu qù?，投射出二元回答“去”qù / “不去”bú qù，排除了旁支语气化。

例7——连动式：话语“他去商店\_\_”Tā qù shāngdiàn \_\_在连动结构的逻辑中召唤一个目的动词：“买东西”mǎi dōngxi、“看朋友”kàn péngyou。

例8——否定与辖域：开头“我不是不\_\_”Wǒ bú shì bù \_\_选择一个意向动词：“想去”xiǎng qù、“喜欢你”xǐhuan nǐ。这一双重否定在下游投射出一个由“只是”zhǐshì或“而是”érshì引导的转折命题。

### 三、汉语的语用预测

语用预测关注最宽泛的单位：叙事图式、会话框架、仪式序列。在这一层面上，预测不再针对一个字或一个命题，而是针对预期的言语行为类型、该行为应当采取的形式，以及它在社会结构中所占据的位置。汉语在此展现出极高密度的文化脚本，使得语用预测具有特殊的结构性，往往超过西方现代语言。

语用预测的第一个轴线涉及问候与告别的仪式脚本。“你好”Nǐ hǎo—

—“你好”Nǐ hǎo 的交流是稳定的。“吃了吗？”Chī le ma? 这一典型的寒暄公式并不期待对饮食的描述，而是期待一个程式化回答：“吃了，你呢？”Chī le, nǐ ne? 同样，分别时“慢走”màn zǒu 召唤“好的，再见”hǎo de, zàijiàn。

第二个轴线是面子（面子，miànzi）管理和仪式性谦虚。面对“你的中文说得真好！”Nǐ de Zhōngwén shuō de zhēn hǎo! 这样的赞美，文化期望的回答不是直接致谢，而是一种谦辞：“哪里哪里，还差得远”Nǎlǐ nǎlǐ, hái chà de yuǎn, 或“过奖了”Guòjiǎng le。

第三个轴线涉及邀请、提供和仪式性拒绝。当主人提出“这是一点小意思，请收下”Zhè shì yìdiǎn xiǎo yìsi, qǐng shōu xià 时，客人应先拒绝：“太客气了，不用了”Tài kèqi le, búyòng le。主人坚持，最终的接受在一两次程式化拒绝后才出现。这种“坚持—拒绝—接受”的模式具有高度可预测性。

第四个轴线涉及餐厅买单。“我来买单”Wǒ lái mǎi dān 召唤“不行不行，我请客”Bù xíng bù xíng, wǒ qǐng kè。这种争相付账的仪式可能持续多个回合。

第五个轴线涉及间接请求。直接请求“给我一杯水”Gěi wǒ yì bēi shuǐ 在大多数社交语境中会被认为是唐突的；预期的形式是“麻烦你，能不

能给我一杯水？”Máfan nǐ, néng bu néng gěi wǒ yì bēi shuǐ? “麻烦”máfan和“能不能”néng bu néng的出现构成礼貌请求的稳定预测标记。

第六个轴线涉及叙事图式。成语往往是故事的凝缩，提到“守株待兔”shǒu zhū dài tù就唤起了一个完整的叙事图式。同样，“塞翁失马”sài wēng shī mǎ唤起道家故事：任何不幸都可能蕴含隐藏的福祀，反之亦然。

第七个轴线涉及引导会话预期后续的句末助词。“吧”ba投射出温和的建议：“我们去吃饭吧”Wǒmen qù chī fàn ba——“好啊”Hǎo a。“呢”ne将问题转向对话者：“我喜欢红色，你呢？”Wǒ xǐhuan hóngsè, nǐ ne? “嘛”ma标记显而易见性：“他是小孩嘛”Tā shì xiǎohái ma。每个句末助词都结构化着下一个话轮。

以下例句在语境中展示了这些预测图式：

例1——同事重逢：“张老师，好久不见！”Zhāng lǎoshī, hǎo jiǔ bú jiàn! 投射出“是啊，最近怎么样？”Shì a, zuìjìn zěnmeyàng?，随后是一个相互恭维和询问家庭状况的阶段。

例2——晚餐邀请：“今晚来我家吃饭吧”Jīnwǎn lái wǒ jiā chī fàn ba首先召唤“不好意思，太麻烦你了”Bù

hǎoyìsi, tài máfan nǐ le, 然后在坚持后才接受。

例3——送礼与收礼：伴随“一点小礼物，不成敬意”Yìdiǎn xiǎo lǐwù, bù chéng jìngyì的赠礼召唤“您太客气了”Nín tài kèqi le, 随后才是接受——通常不会当着赠礼者的面立即打开包装。

例4——道歉与修复：“对不起，我来晚了”Duìbuqǐ, wǒ lái wǎn le投射出一种缓和：“没关系，我也刚到”Méi guānxi, wǒ yě gāng dào, 即使对话者已经等待了很长时间。

例5——电话结束语：序列“好的，就这样”Hǎo de, jiù zhèyàng投射出“好，再联系”Hǎo, zài liánxì, 然后“拜拜”báibái。

例6——市场上的商业谈判：卖方报价“一百块”yì bǎi kuài从买方一侧投射出“太贵了，便宜一点吧”Tài guì le, piányi yìdiǎn ba, 随后是低价还价、拒绝，然后逐步趋同。

例7——亲近者之间的请求：开头“跟你商量一件事”Gēn nǐ shāngliang yí jiàn shì预告了一个请求，其内容将被推迟，先有一段较长的语境铺垫。最终的请求通常由“其实……，不知道……方便不方便”qíshí..., bù zhīdào... fāngbiàn bu fāngbiàn引入。

这三个预测层面——语义、句

法、语用——并非独立的层次：它们持续相互制约，每一层面都限制着其他层面的可能选择。汉语从预测角度看的独特性，与其说在于它们各自孤立地呈现的面貌，不如说在于它们的关联方式。

## 预测性语言模型

上述论述开辟了超越语法描写或普通话教学范畴的视野。汉语预测语法不仅是面向学习者的教学工具，也是一个启发性框架，其成果直接关涉大型语言模型的设计，乃至更广泛的汉语自动处理应用。这种语言学与人工智能之间的交汇值得深入阐明。

当代神经网络架构——无论是Transformer模型、文本扩散模型还是混合系统——归根结底都建立在预测机制之上。它们在每一步估计下一个单位在给定语境下的概率，恰如前述公式 $P(u_n | u_1 \dots u_{n-1})$ 所表达的那样。

因此，模型的质量紧密依赖于其习得的规律与实际支配目标语言的规律之间的契合程度。然而，现有大型模型主要在以英语为主导的语料上训练，倾向于将汉语作为众多语言之一加以处理，而未系统考虑我们所识别的三个结构特征：概率性分词、区别性声调和高度文化语境性。

这种同质化产生的模型虽然可以运作，却难以准确再现普通话特有的预测概貌，在生成时产出形式上正确但语用上失当的话语。

汉语预测语法对此类模型设计的第一项贡献涉及分词。承袭自英语处理的盲目子词切分（如字节对编码）会随意切割复合字符，破坏构成汉字结构的偏旁—声旁关联。以预测语言学为指导的分词将整合偏旁维度（部首）、双音节词的凝聚度以及成语的典型性。

第二项贡献涉及助词和关联结构的处理。体貌助词（了、过、着）、语气助词（吗、呢、吧）和关联对是长距离预测算子。一个适配汉语的预测模型应当将这些标记明确识别为分支点，从这些点出发，可能的后续空间重新组织。

第三项贡献涉及语用和文化维度。一个面对赞美“你的中文说得真好”只生成简单“谢谢”的语言模型，在技术上可能正确，但在文化上是不恰当的。它缺少了中国说话者期望的仪式性谦辞（哪里哪里，还差得远呢）。

这种“文化适配”的要求不应与简单的表层本地化相混淆。更深层地说，要认识到语言从来不是文化中立的编码，而是一个编码着世界表征、社会等级和互动模式的共享期望系统。面子管理、“礼”与“仁”之间的关系、对谦虚的推崇、对含蓄和

间接的偏好——这些都不是附加于中性语法之上的语用装饰，而是汉语可预测性的内在组成部分。

在更具前瞻性的层面上，预测语言学与人工智能的结合勾勒出可被称为“语言学上精确”的新一代模型的轮廓。“精确”，首先是因为这些模型将承认所处理语言的类型学特殊性。“精确”，其次是因为它们将在大规模分布学习之外整合显性的语言学知识作为归纳偏置或控制模块。“精确”，最后是因为它们可以在不仅仅是形式性的、而是文化上定位的标准上接受评估。

这样的取向将有助于减少以英语为中心的模型的语言和文化霸权，保护各语言特有的预测概貌的多样性，并使人工智能成为文化调解的工具，而非同质化的载体。

## 结论

综上所述，预测语言学远不止是一个简单的统计分析框架：它为重新思考现代汉语语法提供了一个富有成效的理论矩阵。传统描写语法通过规则目录和结构清单来进行分析，而预测视角则邀请我们将目光转向使某一接续成为可能的条件，转向组织说话者和听话者期望的机制。预测语言学并不取代既有方法，而是通过赋予它们一个显性的

概率基础来重新奠基它们。

我们依次考察的三个分析层面——语义、句法和语用——的关联并非仅仅是一种陈述上的便利。它反映了汉语语言能力的内在架构。

在语义层面，我们看到预测在最精细的尺度——字和词——上运作，通过双音节搭配的密度、偏旁生产力、成语的结晶化以及量词的制约力来实现。

在句法层面，预测在短语和句子的尺度上展开，调动了如“把”bǎ和“被”bèi这样的结构先导、“因为……所以……”“虽然……但是……”“不但……而且……”等关联对、体貌助词“了”“过”“着”，以及作为主导信息图式的话题—述题结构。

在语用层面，预测采取了对仪式序列、预期言语行为和凝缩叙事图式的预期形式，其中体现了面子（面子，miànzi）管理、间接拒绝策略、程式化谦虚表演，以及如“起承转合”（qǐ-chéng-zhuǎn-hé）这样的话语展开经典图式。

这三个层面远非相互隔绝：它们持续相互制约。句末助词的选择预设了语用预期，话题的识别预设了句法结构，偏旁的辨认则引导了语义预测。

然而，必须强调的是，所揭示的预测机制绝非装饰性的遗存、类型

学上的珍奇或可选的修饰。恰恰相反，它们构成了语言的核心运作机制和可理解性的基础。一个不了解仪式性拒绝脚本或面对赞美时谦虚惯例的普通话使用者，从字面意义上说，将无法正确预测其交流的后续发展。在这个意义上，预测语言学确认了一个古老的直觉：对汉语的掌握不可分割地就是对一种文化的掌握。

我们所勾勒的预测语法面向双重受众。对于汉语学习者，它提供了一种渐进的、语境化的习得方法。学习者不是将语法作为一套需要记忆的规定来学习，而是被邀请发展对规律的敏感性，内化期望视域，训练自己预测即将到来的内容。

对于语言模型设计者，预测语法提供了一份为构建语言学上精确、文化上适切的汉语人工智能而需建模的规律的明确图谱。一套专门的预测语法可以向架构中注入有针对性的归纳偏置，引导学习朝向文化上显著的配置，并在不仅仅是统计合理性而且是语用恰当性的标准上评估产出。

因此，我们所勾勒的汉语预测语法不是一个终点，而是一个起点。它呼唤实证发展、理论延伸，并勾勒一种新型应用语言学的蓝图，其中语言教学与人工智能设计不再是两个平行领域，而是被认识为同一项目的两面：理解、形式化和传

递每种语言中使意义成为可能的东西。汉语凭借其向主导范式发出的挑战，不仅仅是众多研究对象之一：在预测语言学领域，它是一个特殊的实验室，一个真正多语言且文化敏感的人工智能之可能性在此得以验证的场所。

## 参考文献

- Chen, T., & Zhang, D. (2023). Different aspects of vocabulary depth knowledge in L2 Chinese reading comprehension: Comparing higher-and lower-proficiency readers. *Foreign Language Annals*, 56(3), 786-806.
- Chen, X., Wang, S., & Hartsuiker, R. J. (2022). Error-based structure prediction in language comprehension: Evidence from verb bias effects in a visual-world structural priming paradigm for Mandarin Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(1), 60.
- Chen, Y., Li, L., Wang, M., & Wang, R. (2022). Which cognitive factors predict L2 grammar learning: Cognitive control, statistical learning, working memory, or attention?. *Frontiers in psychology*, 13, 943988.
- Fang, H., Jiang, G., & Li, D. (2022). Sentiment analysis based on Chinese BERT and fused deep neural networks for sentence-level Chinese e-commerce product reviews. *Systems Science & Control Engineering*, 10(1), 802-810.
- Guidère, M. (2015). *La Linguistique prédictive : de la cognition à l'action*. Paris : Éditions L'Harmattan. Collection Traductologie.
- Guidère, M. (2025). *La linguistique prédictive : concepts, méthodes et applications*. *Journal of Applied Research in Human & Social Sciences, JARHSS 2025*. Vol. 1. N°1. 72-89. ISSN 2677-7193.
- Hao, Y., Wang, X., Bin, S., Yang, Q., & Liu, H. (2024). How syntactic complexity indices predict Chinese L2 writing quality: An analysis of unified dependency syntactically-annotated corpus. *Assessing Writing*, 61, 100847.
- Huang, Z., Feng, C., & Qu, Q. (2023). Predicting coarse-grained semantic features in language comprehension: evidence from ERP representational similarity analysis and Chinese classifier. *Cerebral Cortex*, 33(13), 8312-8320.
- Li, L., Li, J., Wang, H., & Nie, J. (2024). Application of the transformer model algorithm in chinese word sense disambiguation: a case study in chinese language. *Scientific reports*, 14(1), 6320.
- Li, Y., Qin, S., Ye, J., Huang, H., Li, Y., Guo, S. Y., ... & Yu, P. S.

- (2025, July). Rethinking the roles of large language models in Chinese grammatical error correction. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)* (pp. 553-567).
- Liu, G. L. (2024). Modeling L2 motivation change and its predictive effects on learning behaviors in the extramural digital context: A quantitative investigation in China. *Linguistics Vanguard*, 10(1), 335-343.
- Liu, M., & Dong, L. (2023). A longitudinal study of foreign language reading anxiety and its predicting effects on Chinese university students' reading English performance. *The Asia-Pacific Education Researcher*, 32(6), 783-791.
- Lu, X., & Wu, J. (2022). Noun-phrase complexity measures in Chinese and their relationship to L2 Chinese writing quality: A comparison with topic-comment-unit-based measures. *The Modern Language Journal*, 106(1), 267-283.
- Peng, Y., Sun, J., Quan, J., Wang, Y., Lv, C., & Zhang, H. (2023). Predicting Chinese EFL learners' human-rated writing quality in argumentative writing through multidimensional computational indices of lexical complexity. *Assessing Writing*, 56, 100722.
- Qin, C., Zhang, R., & Xiao, Y. (2022). A questionnaire-based validation of metacognitive strategies in writing and their predictive effects on the writing performance of English as foreign language student writers. *Frontiers in Psychology*, 13, 1071907.
- So, W. C., & Song, X. K. (2023). Whose gestures are more predictive of expressive language abilities among Chinese-speaking children with autism? A comparison of caregivers' and children's gestures. *Journal of Autism and Developmental Disorders*, 53(9), 3449-3459.
- Sun, P. P. (2023). The influence of cognitive, affective, and sociocultural individual differences on L2 Chinese speech performance: a multiple indicators multiple causes approach. *Journal of Psycholinguistic Research*, 52(1), 217-239.
- Wang, Q., Liu, W., Wang, X., Chen, X., Chen, G., & Wu, Q. (2022). A spatial-temporal graph model for pronunciation feature prediction of Chinese poetry. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12), 10294-10308.
- Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology*, 14, 1249991.
- Wei, W., Huang, Z., Feng, C., & Qu, Q. (2023). Predicting phonological information in language comprehension: evidence from ERP representational similarity analysis and Chinese idioms. *Cerebral Cortex*, 33(15), 9367-9375.
- Wu, X., & Wu, Y. (2022, December). From spelling to grammar: A new framework for Chinese grammatical error correction. In *Findings of the Association for Computational Linguistics: EMNLP 2022* (pp. 889-902).
- Xu, M., Wei, Z., & Wu, J. (2022). How emotional communication happens in social media: Predicting "Arousal-Homophily-Echo" emotional communication with multi-dimensional features. *Telematics and Informatics Reports*, 8, 100019.
- Xu, X., Ji, C., Li, T., & Pickering, M. J. (2025). The prediction of segmental and tonal information in Mandarin Chinese: An eye-tracking investigation. *Language, Cognition and Neuroscience*, 40(1), 56-70.
- Yang, H., He, Y., Bu, X., Xu, H., & Guo, W. (2023). Automatic essay evaluation technologies in Chinese writing—a systematic literature review. *Applied Sciences*, 13(19), 10737.
- Ye, J., Li, Y., Ma, S., Xie, R., Wu, W., & Zheng, H. T. (2022). Focus is what you need for Chinese grammatical error correction. *arXiv preprint arXiv:2210.12692*.
- Zhang, J., & Zhang, L. J. (2023). Examining the relationship between English as a foreign language learners' cognitive abilities and L2 grit in predicting their writing performance. *Learning and Instruction*, 88, 101808.
- Zhang, R., & Chen, Y. (2026). What can multi-factors contribute to Chinese EFL learners' implicit L2 knowledge?. *International Review of Applied Linguistics in Language Teaching*, 64(1), 173-195.
- Zhao, X., Lan, G., & Zhang, H. (2024). The predictive effect of language achievement on multiple emotions in languages other than English: Validating a distal mediation model based on the control-value theory. *Applied Linguistics Review*, 15(6), 2713-2732.
- Zhao, Y., & Wu, X. (2022). Predicting reading fluency in Chinese deaf and hard of hearing students: Contributions of character recognition, expressive vocabulary, and syntactic awareness. *American Annals of the Deaf*, 166(5), 663-680.
- Zheng, R., Zhang, M., Guo, T., Guasch, M., & Ferré, P. (2023). Emotional prototypicality ratings for 636 Chinese words: A database of Chinese words with affective information. *Journal of Psycholinguistic Research*, 52(6), 2775-2792.
- Zhou, W., & Gao, B. (2025). Construction and application of English-Chinese multimodal emotional corpus based on artificial intelligence. *International Journal of Human-Computer Interaction*, 41(3), 1782-1793.