中文阅读中的信息密度与认知资源动态分配研究

梁玉豪[1,2]，钟晓路[1]，杨泉[2*]，
1.贵州大学国际教育学院，贵州贵阳
2.北京师范大学国际中文教育学院，北京
yhliang1@gzu.edu.cn 448281784@qq.com yangquan@bnu.edu.cn

摘要

本研究通过分析北京句子语料库的眼动数据，运用混合效应模型和贝叶斯分析方法，系统考察了信息密度在汉语阅读过程中的表现及其与视觉复杂度因素的交互作用。研究结果表明，信息密度对注视时长具有显著正向预测作用，信息密度越高的词汇，受试者的注视时间越长，这与预测编码理论中“预测误差”增加导致加工负荷增加的假设一致；同时，信息密度在跳读行为分析中显示出显著负向预测作用，表明信息密度较高的词越不容易被跳读，支持了读者依据信息分布动态分配注意力的“调节假设”。研究还发现了汉语阅读的语言特异性表现：首先，词长效应在中文中呈现与拼音文字不同的模式，长词在中文中更易被跳读；其次，视觉复杂度与语言预测性之间存在非线性交互，支持了“语言特定性假设”。基于这些发现，本研究提出了中文阅读的“双通道加工模型”，即语言预测（信息密度）与视觉编码（笔画数、词长）共同调节认知资源的动态分配，这一理论框架不仅解释了中文阅读的特异性机制，也为跨语言认知加工研究提供了新视角。

关键词： 信息密度；认知资源；认知负荷

A Study of Information Density and Dynamic Allocation of Cognitive Resources in Chinese Reading

Yuhao Liang[1,2], Xiaolu Zhong[1], Quan Yang[2*],
1.College of International Education, Guizhou University, Guiyang, China
2.School of International Chinese Language Education,Beijing Normal University, Beijing, China
yhliang1@gzu.edu.cn 448281784@qq.com yangquan@bnu.edu.cn

Abstract

This study utilized eye-tracking data from 60 participants reading 150 Chinese sentences in the Beijing Sentence Corpus (BSC), employing mixed-effects models and Bayesian analysis methods to systematically examine the role of information density in Chinese reading processes and its interaction with visual complexity factors. The results indicate that information density has a significant positive predictive effect on fixation duration, with higher information density words receiving longer fixation times from participants, consistent with the hypothesis in predictive coding theory that increased predictive error leads to increased processing load. Additionally, information

*通讯作者

density exhibits a significant negative predictive effect in skimming behavior analysis, suggesting that words with higher information density are less likely to be skimmed, supporting the "modulation hypothesis" that readers dynamically allocate attention based on information distribution. The study also identified language-specific characteristics of Chinese reading: first, the word length effect exhibits a different pattern in Chinese compared to phonetic scripts, with longer words being more prone to skimming in Chinese; second, there is a nonlinear interaction between visual complexity and linguistic predictability, supporting the "language specificity hypothesis." Based on these findings, this study proposes a "dual-channel processing model" for Chinese reading, where language predictability (information density) and visual encoding (stroke count, word length, etc.) jointly regulate the dynamic allocation of cognitive resources. This theoretical framework not only explains the specific mechanisms of Chinese reading but also provides new insights for cross-linguistic cognitive processing research.

Keywords: Information Density , Cognitive resources , Cognitive Load

1 引言

语言作为人类认知的核心载体，其信息传递效率与认知资源分配的动态平衡始终是计算语言学与心理语言学共同关注的焦点。人类语言理解本质上是一种高度动态的认知资源分配过程，并非被动接受信息，而是通过预测—验证机制对语言输入的信息密度（Information Density）进行实时评估与资源调度 (Gibbon, 2000; Frank and Jaeger, 2008)。在实时语言处理中，认知系统必须以毫秒级精度对连续语言信号进行解析，而这一过程受到认知容量的严格约束。预测编码理论（Predictive Coding Theory）为理解这一机制提供了重要框架，指出大脑通过持续更新内部模型来优化认知资源配置，当信息密度超过认知系统的瞬时处理阈值时，触发工作记忆重新加载或预测机制的适应性调整 (Hale, 2001; Friston, 2010)。这一过程在时间维度上与语言信息流形成精细互动，而眼动追踪技术凭借其毫秒级时间分辨率，为观测这种动态互动提供了独特窗口 (Rayner, 2009)。

大量研究已证明信息密度能有效预测阅读过程中眼动模式的变化（如注视时间延长、跳读率降低），为预测编码理论提出的认知资源分配"自适应性"假设提供了强有力支持 (Frank et al., 2015; Rayner et al., 2004)。然而，现有研究主要集中在拼音文字系统，对于汉语这类非拼音文字系统中资源分配机制的探索仍然有限。汉语作为非拼音文字系统，具有区别于拼音文字的独特特征，为认知资源的动态调度机制提出了新的理论挑战 (Li et al., 2014)。汉语缺乏明确的词边界标记，迫使读者在实时加工中主动完成词汇切分，这一过程不仅依赖语义句法预测，更依赖视觉空间线索和词长分布规律，重构了信息密度的作用路径。同时，作为孤立语，汉语语法关系主要依赖语序和虚词，句法结构相对灵活且上下文依赖性强，要求更强的句法预测与实时整合，影响词汇在视觉空间序列中的预期位置及整体性感知 (Wei et al., 2018)。此外，汉字形-义-音联结方式多样，复合词语义透明度差异大，理解需要基于语境的语义预测，其内部结构的可分析性和整体视觉形态也直接影响识别效率。这些特征表明，在汉语阅读中，信息密度驱动的预测机制如何与视觉特征驱动的加工需求相互作用，以及这种双通道信息如何共同调节认知资源分配，目前尚缺乏系统理解。

基于上述理论背景，我们认为中文阅读的认知资源分配必然是语言预测（信息密度、可预测性、句法预测）与视觉空间编码（笔画数、词长、整体轮廓、空间布局线索）协同作用的结果。单一的"信息密度通道"模型难以充分解释汉语阅读中诸如"词长悖论"（长词更易跳读）、词边界切分负荷以及高语义密度下的高效加工等现象。因此，本研究聚焦视觉模态下的文本阅读，通过眼动追踪技术捕捉实时认知资源分配，基于大规模标准化眼动数据库，构建词汇级认知负荷动态模型。我们尝试构建整合"语言预测通道"与"视觉空间编码通道"的双通道模型，联合注视时长与跳读率等眼动指标，探索信息密度对中文阅读中认知资源分配的预测力，旨在提供更全面、更具生态效度的认知解释框架。

2 相关研究

2.1 信息密度与惊异度

语言信息密度的研究根植于信息论的基本原理，尤其是Shannon (1948)提出的信息量（information content）概念在语言学中的跨学科应用。他指出信息的本质在于对不确定性的消除，而语言作为信息传递的媒介，其结构和表达方式也可以用概率和熵的框架进行量化分析。基于这一理论框架Hale (2001)提出“惊异度（Surprisal）”概念，用以量化听者或读者在增量式语言加工中对即将到来的词项的预测难度，这一量值反映了语言单位的信息量，并据此推测理解该单位时所需的认知资源投入。换言之，惊异度越高，代表该词项越不可预测，其处理成本也随之上升。此后的大量研究进一步证实，惊异度在很大程度上反映了预测与实际输入之间的差异，并与阅读时的注视时长和脑电成分（如N400）高度相关 (Smith and Levy, 2013; Futrell et al., 2021)。

在认知加工框架中，惊异度不仅被视为处理难度的定量指标，还可以整合入对认知资源分配的建模之中，捕捉歧义消解、期望更新等动态过程中的资源负担 (Levy, 2008)。语言接受者在语境中不断生成对后续语言输入的概率性预测，因此语言的处理难度不仅由其形式决定，更取决于其在上下文中的预测概率。信息密度理论由此奠定基础：语言理解是一种动态分配认知资源的过程，信息分布的稀疏与密集直接影响加工效率。在这一背景下，学界进一步发展出“均匀信息密度假说”（Uniform Information Density, UID），认为语言使用者倾向于在语句中平衡信息分布，从而避免信息过于集中或稀疏，以优化沟通效率和认知可负担性 (Levy and Jaeger, 2007; Genzel and Charniak, 2023)。UID 假说预言，信息密度（如surprisal）突增会导致处理负荷非线性增长，这一预测已在阅读时间、语法选择和语言生成的多个研究中得到支持 (Meister et al., 2021)。因此，惊异度不仅作为理论上的信息度量指标，其对眼动行为、神经响应等多模态数据的良好预测力也使其成为语言加工认知负荷的重要操作化工具。

2.2 眼动行为与认知资源分配

在心理语言学中，眼动追踪技术被广泛用于探索认知资源在语言处理中的分配方式。通过记录眼睛在阅读或听语言时的运动，研究者能够深入了解大脑如何处理语言信息，以及在不同情境下如何分配注意力。眼动数据中的关键指标包括注视时长（Fixation Duration, FD）、凝视时长（Gaze Duration, GD）、总阅读时间（Total Reading Time, TRT）以及跳读率（Skipping Rate）等 (Rayner, 2009)。这些指标分别反映了词汇加工的不同阶段，如词汇识别、语义整合和高级句法处理等，以及认知负荷的高低和处理效率。

注视时长直接反映了处理某个词或区域所需的认知努力，可以分为早期指标（如首次注视时长）和晚期指标（如总阅读时间）。早期指标通常反映自动化的词语识别过程，而晚期指标则与更策略性的处理（如重新分析或整合）相关，通常认为较长的注视时长表示较高的认知负荷 (Ke et al., 2024)。跳读率是指读者在阅读时跳过某些词而不直接注视，通常发生在词语可以通过周边视觉预处理或在上下文中高度可预测时。跳读是早期处理指标，一般与周边视觉预处理（parafoveal preview）相关，读者在注视当前词时已经开始处理下一个词的信息，反映了较低的认知负荷和高效的词语识别 (Staub and Rayner, 2007)。在拉丁字母语言的研究中，多项研究已经确立了影响阅读眼动的关键因素，包括词频、词长、词的可预测性和模糊性等 (Kliegl et al., 2004; Staub and Benatar, 2013)。其中，词频效应是最为稳健的发现之一，表现为高频词的注视时间显著短于低频词 (Inhoff and Rayner, 1986)。词长效应则表现为长词的注视时间长于短词，且短词更容易被跳读 (Brysbaert et al., 2005)。这些效应反映了读者在阅读过程中的注意资源分配策略，为理解阅读的认知机制提供了重要线索。

2.3 汉语阅读的视觉特性与加工

汉语作为典型的孤立语与表意文字系统，其阅读加工的认知资源分配机制呈现出显著的语言特异性。首先，汉语缺乏显性词边界标记的特征迫使读者在实时加工中主动完成词汇切分 (Li et al., 2009)，这一过程不仅涉及语义和句法预测，还需额外投入认知资源进行词边界判定，从而可能重构信息密度的作用路径。Yan (2010)提出的“动态目标理论”揭示，中文读者能够基于旁注视信息灵活切换字词层级的注视目标，这种动态调整机制暗示信息密度驱动的资源分配需与词汇切分过程形成协同。更为特殊的是，汉语的词长由汉字数量（而非字母数）构成，通常以1-4字为基本单元，其加工效应表现出与拉丁语系相反的“词长悖论”：部分长词（如四字成

语）因其高语义完整性反而呈现加工优势 (Liversedge et al., 2016; Zang et al., 2011)。这种特性提示，汉语阅读中的认知资源调度不仅受信息密度的调控，更与词汇的视觉-语义整合效率密切相关，构成区别于拼音文字的结构-功能双维度资源分配模式。

其次，汉字的视觉复杂性对认知资源分配具有独特的调节作用。作为表意文字系统的一部分，汉字具有方块字布局、复杂的笔画结构和高度整合的部件配置，这些视觉特征在加工过程中占据显著的感知资源。已有研究指出，笔画数可作为视觉复杂度的有效指标，笔画越多的汉字通常需要更长的注视时间以完成视觉识别 (Ma and Li, 2015; Jiang et al., 2020)。尽管书写与识别的认知路径存在差异，但高笔画数导致的视觉轮廓模糊性和特征提取负荷仍会延长早期加工阶段（如首次注视时长）。最新研究进一步发现，笔画数效应在词汇学习的不同阶段始终存在，说明其作用主要集中于词汇加工的低层级视觉识别过程，并不因语言经验的积累而减弱 (Liang et al., 2024)。这一发现支持“视觉限制性假说”，强调视觉特征在中文阅读中的基础性调控作用。

同时，由于汉字的形义紧密对应关系，其视觉特征往往与语义激活过程高度耦合，这使得视觉加工和语义提取过程在时间上更加重叠 (Perfetti and Tan, 1999)。因此，中文阅读中存在一种融合视觉识别与语言预测的信息处理模式，即“双通道加工”机制 (Li et al., 2011)，对认知资源的动态分配提出了更高要求。与此相应，在眼动行为上也表现出不同于字母语言的特点。例如，中文阅读中词长与跳读率呈现悖论性正相关，即较长的词反而更可能被跳过，这与英语中长词更难跳过的负相关模式形成鲜明对比 (Rayner, 1998)。这种差异可能源于语言结构和词边界识别方式的根本不同。中文缺乏显式的词间空格，使得词的整体视觉形状、词频以及句法语境成为判断词边界的重要线索 (Li et al., 2009)。在此基础上，长词往往具有更稳定的视觉特征和语义整体性，从而更容易在周边视野中被识别为整体加工单位，被眼动系统“略读”或跳过 (Yan et al., 2010)。这一现象进一步印证了中文阅读在资源分配策略上对视觉线索和语言预测的高度依赖。此外，中文的语义密度较高，单字即可携带丰富信息，这可能影响信息密度的计算和认知表现 (Liversedge et al., 2016)。

综上所述，虽然现有研究已揭示了中文阅读的若干特性，但仍缺乏系统整合信息密度理论的研究框架，特别是对信息密度如何与中文特有的视觉和语言特性交互影响认知资源分配的理解仍然有限。信息密度理论虽然为理解认知资源分配提供了重要框架，但其在中文阅读中的适用性及与中文特有视觉特征的交互作用仍需系统研究。据此，本研究聚焦以下三个核心问题：

Q1：信息密度如何影响中文阅读中的认知资源分配?

Q2：信息密度与词汇属性（可预测性、词频、词长、笔画数）如何交互影响认知资源分配?

Q3：是否可以构建基于信息密度的认知资源分配预测模型，有效模拟中文阅读过程中的眼动行为?

通过回答这些问题，本研究试图揭示中文阅读中基于信息密度的认知资源分配机制，并提出整合语言预测（信息密度）和视觉编码（笔画数、词长）的“双通道资源分配模型”，以理解中文阅读中的认知资源动态分配规律。这不仅能验证信息密度理论在跨语言语境下的普适性和特异性，也为基于认知机制的语言处理技术提供理论基础。

3 数据与方法

3.1 数据来源

本文所选取的眼动数据库为北京语句语料库（Beijing Sentence Corpus，BSC），这是首个公开发布的大规模标准化中文阅读眼动数据集 (Yan et al., 2021; Pan et al., 2022)。该数据库收录了60名中文母语者在自然阅读条件下的完整眼动轨迹，精确记录了注视时间、注视位置、扫视方向等多维指标。在语料特征方面，BSC的句子材料主要选自《人民日报》，经专业语言学家修改处理，移除了政治敏感内容和词边界歧义因素。句子长度控制在7-15个词（平均11.2词），对应15-25个汉字（平均21字），保证了阅读材料的自然性和生态效度。在词汇分布方面，BSC包含了702个名词、368个动词、183个虚词等多种词类。从词长分布看，1-4字词的占比分别为20.6%、73.7%、4.2%和1.5%，反映了中文词汇以双字词为主的特点。词汇复杂度方面，收录词汇的笔画数范围为2-42画（平均15.6笔）。词频信息基于《现代汉语词频词典》和SUBTLEX-CH数据库计算，平均词频约为5,000次/百万词，确保了词汇样本的代表性和平衡性。此外，BSC还构建了大规模的词汇可预测性规范库。基于148名参与者完成的Cloze测试

（完形填空任务），该数据库为每个词提供了标准化的可预测性评分（logit值）。这些经过严格控制和规范化处理的语言学特征与眼动指标，为探究中文信息密度与认知资源分配的动态关系提供了独特而可靠的数据基础。本研究所选用的部分指标数据如下：

指标	含义
注视时间（dur）	注视持续时间（单位：毫秒），反映词汇加工难度。
跳读率（skip）	根据出视扫视幅度（Amplitude Outgoing，ao）是否大于词长计算二值变量（1=跳读，0=注视），出视扫视幅度是指从当前注视点到下一注视点的距离（单位：字符数）。
可预测性（p）	通过Cloze测试计算的logit值（范围：负值表示低预测性，正值表示高预测性），并解决概率分布偏态问题。
词长（l）	以字符数的倒数表示（如2字符词l=0.5，用于统计模型中的标准化）。
词频（f）	基于语料库的对数转换值（如log10（词频+1））。
笔画数（i）	词的总笔画数，也即视觉复杂度。

Table 1: 北京语句语料库（BSC）部分指标与含义

3.2 信息密度的计算

信息密度指语言单位所承载的信息量在时间或结构上的分布状况，是评估语言加工负荷的核心指标之一，其计算通常基于词汇的惊异度。惊异度的数学定义如下：：

$$Surprisal(w_i) = -\log_2 P(w_i \mid context)$$

其中，w_i 表示目标词，$context$ 为该词的上下文（如前文或双向上下文），$P(w_i \mid context)$ 是语言模型对该词在该上下文中出现的条件概率。假设有这样一个句子：”今天天气真好”，我们通过语言模型或语料库统计得到的条件概率，计算每个词的信息量，最后计算以字为单位的信息密度。其计算示意图如下：

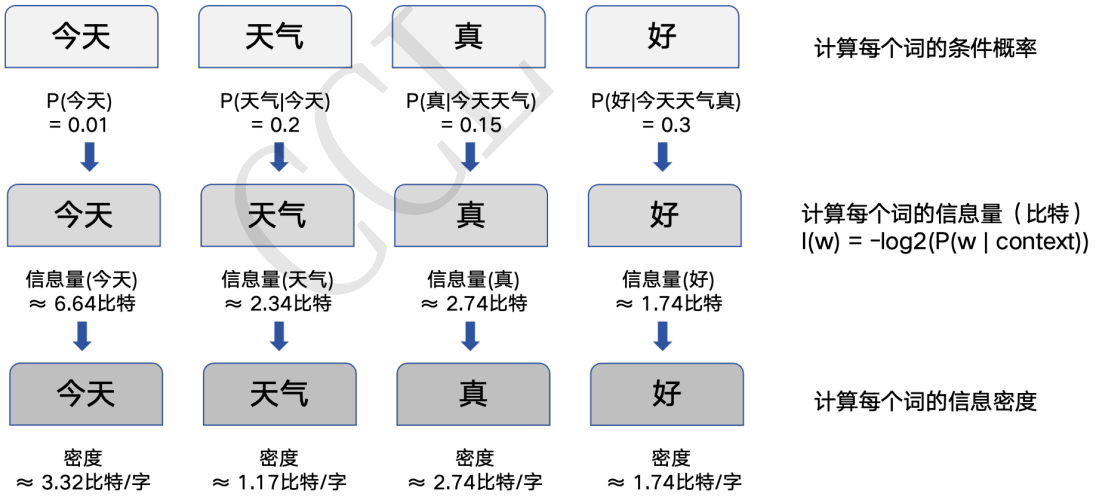

Figure 1: 信息密度计算过程

早期的信息密度研究通常依赖n-gram模型获取上下文概率，但这类模型对长距离依赖和复杂句法结构的建模能力有限。因其只能捕捉有限的上下文信息（通常为2-5个词），难以处理长距离依赖关系，且该模型基于独立性假设，无法充分建模词汇间的复杂语义关系。随着深度神经网络的发展，基于Transformer的预训练语言模型（如BERT、GPT）因其强大的上下文建模能力，逐渐成为计算惊异度和信息密度的一种重要方法 (Wilcox et al., 2020)，许多研究表明BERT计算惊异度对人类处理中的结构异常、语义预测以及神经电生理反应 (Huber et

al., 2024)提供高度一致的预测，显示其在模拟信息密度分布上的有效性。因此，在本研究中，我们使用中文BERT模型来计算词汇的惊异度。BERT是基于掩码语言模型（Masked Language Model, MLM）的结构，我们采用如下步骤将BERT应用于Surprisal计算：对于每个目标词w_i，首先构造包含该词的完整原始句子$S = w_1 w_2 \ldots w_n$，然后将目标词w_i 替换为`[MASK]` token，形成masked 输入序列S_{mask}。接着，使用BERT 模型预测`[MASK]` 位置上的词分布，并从中提取该位置上对应于w_i 的预测概率$P(w_i \mid S_{mask})$，最后计算其Surprisal 值为$-\log_2 P(w_i \mid S_{mask})$。借助BERT 的双向上下文建模能力，我们能够获得更精确的词项条件概率估计。

4 分析与结果

4.1 数据预处理

在进行正式分析之前，我们首先对原始数据进行了预处理。预处理主要包括数据合并、对数变换、分组与标准化等步骤，重点解决多源数据整合、变量尺度差异与模型假设验证三个核心问题。首先，将眼动仪记录的注视时间数据（包含参与者编号、句子编号、词位置等关键变量）与信息密度、词频、词长、笔画数、可预测性等变量，通过句子-词汇位置的主键进行精准匹配，形成初始数据集（$n = 91{,}265$）。其次，通过三阶段数据处理：（1）删除关键变量存在缺失的观测（$n = 217$，占总样本0.24%），缺失机制检验（Little's MCAR 检验：$\chi^2 = 32.15$, $p = 0.18$）表明缺失为完全随机；（2）对右偏态的注视时间变量（偏度= 1.83）实施自然对数变换，经Box-Cox 验证（$\lambda = 0.12$）确认其有效性，变换后的残差通过正态性检验（Shapiro-Wilk，$p = 0.12$）；（3）对五个连续预测变量进行z-score 标准化，消除词频（量级范围$10^{-6} \sim 10^{3}$）与信息密度（量级0.1 ∼ 2.4）等变量的尺度差异。

同时，为确保预测变量之间的多重共线性问题不会影响分析结果，我们计算了预测变量之间的相关系数矩阵。共线性诊断显示，预测变量间相关系数均低于0.7的临界值 (Dormann et al., 2013)，其中词长与笔画数的相关性最高（r=0.687），但方差膨胀因子（VIF=2.83）仍低于经验阈值5。变量标准化后构建的最终数据集包含91048个有效观测，共20个变量。最后，我们通过残差诊断（Q-Q图与Levene检验p=0.24）验证了线性混合模型的正态性、独立性与方差齐性假设，为后续分析提供了可靠的数据基础。

变量	信息密度	可预测性	词频	词长	笔画数
信息密度	1.000000				
可预测性	-0.174345	1.000000			
词频	-0.413753	0.399957	1.000000		
词长	-0.517122	0.371281	0.687707	1.000000	
笔画数	0.330845	-0.235398	-0.497945	-0.584639	1.000000

Table 2: 预测变量相关系数矩阵

4.2 注视时间模型

由于眼动数据具有嵌套结构（词语嵌套在句子中，句子和词语又被不同被试阅读），因此数据点之间存在非独立性。为应对这一问题，我们采用线性混合效应模型（Linear Mixed-Effects Model, LMM）进行分析，该方法能够同时估计固定效应（即我们关注的预测变量的影响）与随机效应（如"被试"和"句子"所带来的变异），从而有效应对嵌套数据结构带来的统计偏误。在本研究中，我们以对数变换后的注视时间（`log_dur`）为因变量，固定效应包括信息密度（`information_density_z`）、可预测性（`p_z`）、词频（`f_z`）、词长（`l_z`）和笔画数（`i_z`），随机效应为被试（`id`）和句子（`sn`）的截距。我们首先构建了一个仅包含主效应的基线模型，其模型形式如下：

$$\texttt{log\_dur} \sim \texttt{information\_density\_z} + \texttt{p\_z} + \texttt{f\_z} + \texttt{l\_z} + \texttt{i\_z} + (1|\texttt{id}) + (1|\texttt{sn})$$

该模型通过限制最大似然法（REML）进行拟合，结果显示"句子"的随机效应方差为0.001，标准差为0.024；"被试"的随机效应方差为0.014，标准差为0.119；残差的方差

为0.133，标准差为0.364。这一结果揭示了不同句子和个体之间的变异程度，同时也表明模型在解释个体与句子间变异方面的能力有限。模型结果如图所示：

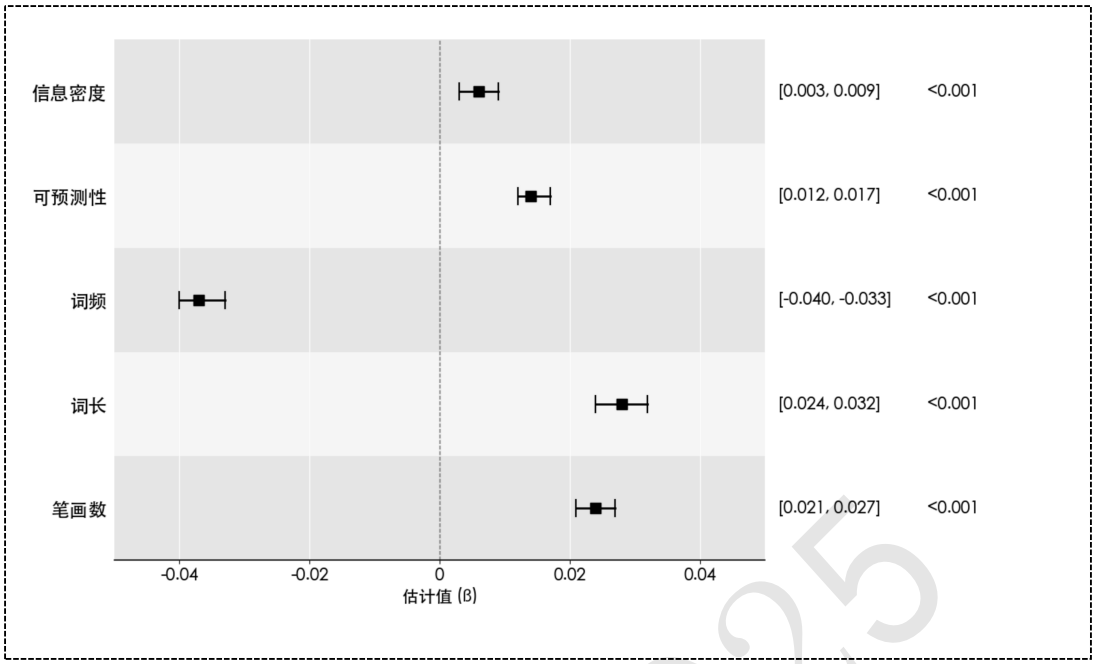

Figure 2: 主效应的模型分析结果

结果显示所有预测变量均对注视时长具有显著影响，其中信息密度的估计系数为0.006（95% CI: [0.003, 0.009]，$p < 0.001$），表明信息密度每增加1 个标准差，对数注视时间增加0.006 个单位，支持预测编码理论，即信息密度越高的词（惊异度越大），注视时间越长，从而验证了惊异度效应在汉语阅读中的存在。长词（$\beta = 0.028$）和高笔画数词（$\beta = 0.024$）也显著延长注视时间，反映出更高的视觉加工负荷，这与视觉编码假设一致。词频的估计系数为$\beta = -0.037$（95% CI: [0.040, 0.033]，$p < 0.001$），呈现显著负相关关系，说明高频词被注视的时间更短。此外，可预测性的估计系数为$\beta = 0.014$（$p < 0.001$），表明可预测性高的词注视时间更长，这与经典研究结果相悖。

为进一步探究信息密度与其他因素的交互作用，我们另外构建了三个包含交互项的模型。第一个模型分析了惊异度与可预测性的交互作用，其模型如下：

$$\texttt{log\_dur} \sim \texttt{information\_density\_z} * \texttt{p\_z} + \texttt{f\_z} + \texttt{l\_z} + \texttt{i\_z} + (1|\texttt{id}) + (1|\texttt{sn})$$

结果显示，信息密度与可预测性的交互效应显著（$\beta = 0.006$，SE $= 0.001$，$p < 0.001$），而信息密度的主效应变得边缘显著（$\beta = 0.003$，SE $= 0.002$，$p = 0.091$）。这一结果揭示出一种更为复杂的认知加工模式：信息密度对注视时间的影响取决于词的可预测性水平。在高可预测性条件下，信息密度对注视时间的正向影响更为明显；而在低可预测性条件下，这种影响则显著减弱。我们进一步通过交互效应的可视化图像验证了这一发现：

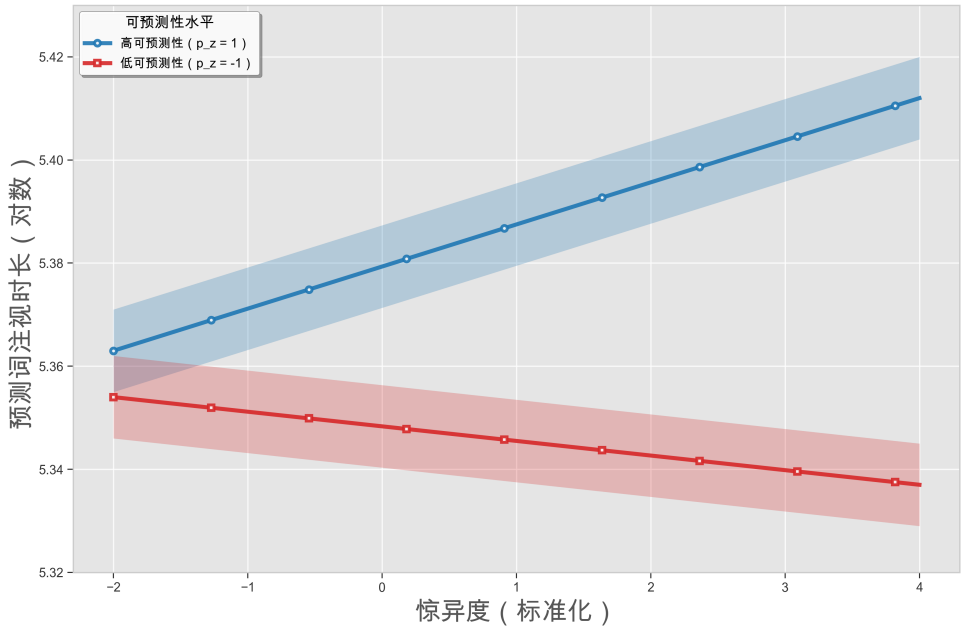

Figure 3: 信息密度与交互效应可视化

将可预测性在±1标准差处分为高低两组，可以清晰观察到高可预测性组中，信息密度增加导致注视时间的增长斜率更陡峭。这暗示读者在处理高可预测性词时，更能有效分配注意资源来应对信息密度变化；而对低可预测性词，则采取相对保守的加工策略，减弱了对信息密度变化的敏感度。

第二个模型检验了信息密度与词长的交互作用：$\texttt{log\_dur} \sim \texttt{information\_density\_z} * \texttt{l\_z} + \texttt{p\_z} + \texttt{f\_z} + \texttt{i\_z} + (1|\texttt{id}) + (1|\texttt{sn})$，结果发现该交互作用同样显著（$\beta = 0.009$，95% CI: [0.006, 0.013]，$p < 0.001$），表明词长越长，惊异度效应越强。第三个模型分析了信息密度与笔画数的交互作用：$\texttt{log\_dur} \sim \texttt{information\_density\_z} * \texttt{i\_z} + \texttt{p\_z} + \texttt{f\_z} + \texttt{l\_z} + (1|\texttt{id}) + (1|\texttt{sn})$，结果显示这一交互作用也达到显著水平（$\beta = -0.009$，95% CI: [0.011, 0.006]，$p < 0.001$），但呈负相关关系，表明笔画数越多，惊异度效应反而越弱。

为确定交互项对模型拟合的贡献，我们通过AIC和BIC指标以及似然比检验比较了不同模型。比较结果显示，加入惊异度与其他因素的交互项显著改善了模型拟合（$\chi^2(1) = 16.433$，$p = 0.0001$）。进一步对三种交互作用模型进行比较，结果表明惊异度与笔画数的交互作用模型的对数似然值最高（−37511.170），AIC值最低（75042.339），提供了最佳拟合效果。这表明在汉语阅读中，汉字的视觉复杂度与惊异度的交互关系对注视时长的影响最为重要，反映了语言认知加工与视觉信息处理的密切互动。

4.3 跳读率模型

跳读变量是基于首次进入时间（ao）和词长（l）来计算的二元跳读变量skip，当ao大于词长的倒数时，认为该词被跳过（skip=1），否则为0，skip是一个0或1的二元变量，其分布符合二项分布（或伯努利分布）。因此，我们选用广义线性混合模型（Generalized Linear Mixed Model, GLMM），以二元跳读变量skip作为因变量。GLMM是LMM的扩展，适用于因变量非正态分布的情况。对于二项分布，默认且最常用的是Logit链接函数，模型预测的是事件（跳读）发生的对数几率（log odds），因此，通过链接函数连接预测变量和因变量的期望值。模型设定为：

$$\texttt{skip} \sim \texttt{information\_density\_z} + \texttt{p\_z} + \texttt{f\_z} + \texttt{l\_z} + \texttt{i\_z} + (1|\texttt{id}) + (1|\texttt{sn})$$

使用pymer4包的Lmer函数，并指定family="binomial"，通过最大似然法拟合，观测值数量、分组情况与模型一同。随机效应同样包含(1|id)和(1|sn)来解释被试和句子的变异性。结果显示sn的方差为0.066，标准差为0.256；id的方差为0.558，标准差为0.747，表明个体之间的变异较大。固定效应的结果则如下：信息密度（$\beta = -0.244$，SE=0.009，$p < 0.001$）、可预测

性（$\beta = 0.095$，SE=0.009，$p < 0.001$）、词频（$\beta = -0.016$，SE=0.012，$p = 0.178$）、词长（$\beta = 0.730$，SE=0.014，$p < 0.001$）和笔画数（$\beta = -0.034$，SE=0.010，$p = 0.001$）。

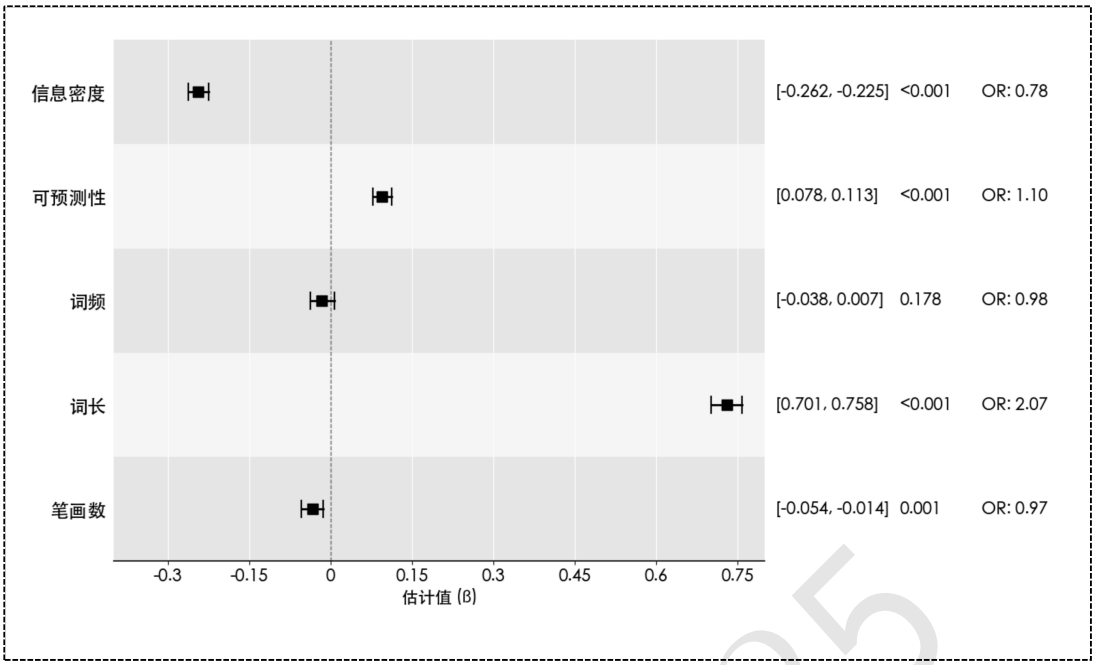

Figure 4: 跳读模型的分析结果

从上述研究我们可以发现，信息密度每增1个标准差，跳读概率降低22%（OR=0.78），符合“高认知负荷抑制跳读”假设。其次，存在词长悖论：长词跳读概率显著更高（OR=2.07），可能与中文无空格下长词更易被整体识别有关 (Yan et al., 2010)。再次，笔画数的增加会显著降低跳读概率（OR=0.97），凸显中文视觉加工特性。

4.4 贝叶斯模型

为模拟读者在阅读过程中的跳读和延长注视决策，并进一步预测具体的注视时间。我们通过贝叶斯方法构建眼动预测模型，贝叶斯分析通过优化参数θ_s（跳读参数）和θ_{dur}（注视时长参数）来最大化模型的预测准确率，构建`simulate_decision`函数模拟读者阅读决策，包括：跳读决策，基于标准化惊异度(Surprisal_z) 与阈值θ_s比较。若Surprisal_z 小于θ_s，预测为跳读(1)，否则为不跳读(0)；延长注视决策，基于多特征(Surprisal_z, f_z, l_z, i_z, p_z) 加权和与阈值θ_{dur}比较。若加权和大于θ_{dur}，预测为延长注视(1)，否则为正常注视(0)。基于标准化信息密度与阈值比较预测跳读行为，结合多特征加权和阈值比较预测延长注视行为。模型使用Optuna库在参数范围(θ_s：-2到2, θ_{dur}：-2到2) 内进行50次试验，寻找使跳读准确率(acc_skip) 和延长注视准确率(acc_ext) 之和最大化的最优参数组合，最终得到最佳参数为$\theta_s = -1.7988400368731112$ 和$\theta_{dur} = 1.9859485527563157$。

最后的结果显示，模型在跳读预测上达到100%的准确率，在延长注视预测上准确率为82.05%，混淆矩阵显示真阳性72519、假阳性4570、假阴性12558、真阴性1694。这些结果表明，该贝叶斯眼动预测模型能有效捕捉阅读过程中跳读和延长注视现象，动态模型很好地学会了根据信息密度判定是否跳过，这与中文的“可预测性驱动跳词”机制高度一致。不过，跳词行为的百分百预测率显示出模型存在一定的“过拟合”风险，因为跳词行为是由词级信息密度主导的“离散决策”（要么跳，要么不跳）。

同时，我们还使用注视时间箱线图从分布特征层面验证模型预测效果，具体来说，比较了句子前15个词位置上实际注视时间（蓝色）和预测注视时间（橙色）的分布。X 轴为词位置(wn)，Y 轴为注视时间（ms）。每个箱体显示了中位数、四分位数范围，线条延伸显示数据范围（排除异常值），点代表异常值。该图详细比较了模型预测与实际数据在每个词位置上的分布特征。

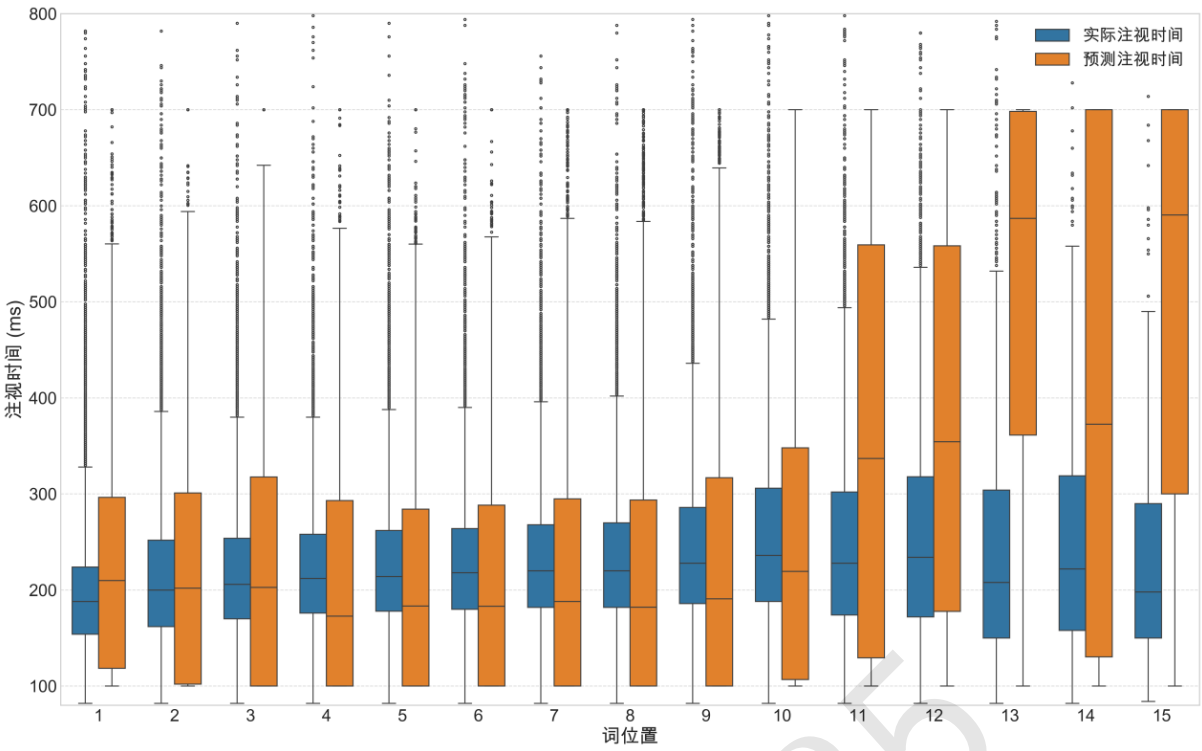

Figure 5: 实际注视时间和预测注视时间分布图

从上图来看，早期位置的句子单词预测效果要好于后期位置。原因可能有二：首先，数据库的句子长度范围是7—15个词，平均为11.2词，这使得后续位置的建模数据样本不足，从注视时间箱线图我们也可以看出，在句子中位置大概处于前10的词语能够展示出良好的预测率，而处于11以后的词语效果偏差较大。其次，延长注视行为是认知负荷的“连续波动”，这会受多个变量共同作用，建模难度更高，因此，在延长注视预测中存在一定误差，后期位置偏差反映了累积的认知负荷效应，这些效应并未被当前的阈值完全捕捉到。不过整体来说，模型展现出良好的拟合能力和预测性能，未来可通过扩展特征集和优化模型结构进一步提升精度。

5 讨论与结论

5.1 信息密度对注视时间的影响机制

首先，信息密度对注视时长具有显著正向预测作用，信息密度越高的词汇，受试者的注视时间越长。这一结果与前人关于信息密度与加工负荷正相关的观点一致 (Hale, 2001; Levy, 2008; Staub, 2015)，支持了语言加工过程中“预测误差”（prediction error）增加会延缓眼动行为的理论假设。在语言理解中，当词汇出现不符合前文语境预测时，意味着该词携带的信息量增加，因而需要更多认知资源进行处理。这与英语 (Frank et al., 2015)和德语 (Kliegl et al., 2006)等研究结果一致，证实了信息密度理论与预测性语言加工理论在跨语言情境下的普适性。然而，信息密度并非孤立地发挥作用，其影响深刻嵌入于与其他词汇特征的复杂交互网络中，并揭示了语言预测通道与视觉编码通道在资源有限条件下的动态互动机制。

其次，信息密度与词长、词频、可预测性等变量之间的交互效应显著。值得注意的是，可预测性的主效应（β=0.014）表面上与经典研究 (Rayner et al., 2004)的发现存在矛盾，但信息密度与可预测性的显著正向交互效应（β=0.006）提供了更深层次的解释：当预测与实际输入不匹配时（高信息密度），认知系统在高可预测性条件下反而需要投入更多资源进行信息的重新整合与预测更新，导致加工成本显著增加。这种调节效应反映了人类语言加工的“资源敏感性”（resource sensitivity）——在认知资源有限条件下，加工系统对可预测语境中出现的突发信息输入（高信息密度）尤为敏感，体现了语言预测通道内部的动态调整。

信息密度与视觉复杂度因素（笔画数、词长）的交互作用揭示了语言预测通道与视觉

编码通道之间深刻的协同与竞争关系。信息密度与笔画数之间存在显著的负向交互（β = -0.009），表明视觉复杂度（高笔画数）的增加削弱了信息密度对注视时间的正向影响。这体现了资源竞争机制：当加工笔画繁多的汉字时，大量认知资源被优先消耗在低层级的视觉特征提取和字形识别（视觉编码通道负荷高），导致可用于高层级语言预测通道（处理信息密度、上下文整合）的资源显著减少，从而削弱了高信息密度本应带来的注视时长延长效应。视觉加工的"瓶颈"制约了语言预测加工的深度。同时，信息密度与词长呈现显著的正向交互（β = 0.009），结合跳读模型中发现的长词更容易被跳读的"词长悖论"（OR=2.07），这反映了视觉编码通道在特定情境下的补偿机制：在缺乏词边界空格的中文中，长词占据更大的视觉空间，其整体轮廓（视觉编码特征）提供了更清晰的感知单元信号，使得读者更容易在副中央凹完成预加工和整体识别。这种视觉整体性部分补偿了可能存在的语言预测不确定性或高信息密度带来的负荷需求，既提高了跳读概率（减少了对中央凹注视的需求），也可能在注视发生时，其视觉整体性为处理高信息密度提供了某种支持。这些发现不仅符合 (Reichle et al., 2013)提出的词汇加工与眼动控制协同关系，更进一步凸显了在汉语环境下，视觉复杂度（如笔画数）与语言预测性之间存在的非线性交互具有语言特异性，支持Perfetti and Harris (2013)提出的"语言特定性假设"，即阅读认知过程会受到特定语言正字法特性的调节。

5.2 信息密度对跳读行为的约束作用

在跳读行为分析中，信息密度显示出显著负向预测作用，即信息密度越高的词越不容易被跳读。该发现进一步强化了信息密度在调节眼动策略中的核心地位。根据阅读研究中的"调节假设"（regulation hypothesis），读者在阅读过程中会依据信息分布动态分配注意力，跳过可预测、低信息量的词，而保留对信息密度较高词的注视，本研究结果为这一机制提供了有力证据。

研究发现词长对跳读的影响具有语言特异性：在中文中，长词更易被跳读（OR = 2.07），这一结果与英语研究中长词通常抑制跳读的结论形成对比 (Drieghe et al., 2005)。传统模型（如E-Z Reader和SWIFT模)通常假设词长与跳读呈负相关关系 (Engbert et al., 2005)。这种语言间差异可能源于书写系统的结构差异，在英文中，词与词之间通过空格进行边界标记，而在中文中，词无显性边界划分，长词往往占据更大的水平空间，更容易在副中央凹（parafoveal）范围内被感知为完整的处理单元，因而更易被预加工并跳过。这一机制与"动态目标模型"的预测一致：中文阅读中，读者会根据边界线索动态选择跳读目标，若边界信息足够明确，则以词为单位跳读，否则以字符为单位 (Yan et al., 2010)。此外，该结果也支持 (2014) 提出的中文阅读中副中央凹信息获取机制的特殊性。需要强调的是，中文词长（以字数计）反映的是视觉单位的空间延展性，而非拼音语言中反映音节复杂度的线性结构。相比之下，拼音文字中的长词通常意味着更长的音素串、更复杂的形态结构，导致解码负荷增加，因而较少被跳读；而在中文中，长词因占据更广的视觉范围而更容易整体识别并被跳过 (Yan et al., 2010)，这种机制上的根本差异凸显了中文阅读加工的独特性。

同样，贝叶斯模型在跳读预测中的高准确率表明，信息密度是中文跳读决策的核心指标之一，支持"经济性原则"——读者倾向于跳过认知负荷低的词汇。然而，模型对延长注视的预测误差揭示了连续认知负荷建模的复杂性，表明需结合工作记忆资源分配等机制 (Reichle et al., 2009)进一步完善理论模型。这一结果为整合眼动控制的认知建模提供了重要启示。

5.3 理论意义与未来研究方向

基于上述发现，本研究提出中文阅读的"双通道加工模型"。该模型的核心在于，认知资源的有限性是理解资源分配的基础。语言预测通道（信息密度、可预测性）与视觉编码通道（笔画数、词长、整体轮廓）并非独立运作，而是在加工过程中进行实时动态交互，主要表现为竞争、协同与补偿，共同调节认知资源的动态分配。竞争机制体现在高视觉复杂度（如笔画数多）会抢占资源用于低层级视觉加工，限制高层级语言预测的资源，削弱信息密度效应。协同机制表现为高可预测性提供的强语境线索能优化资源调配，使读者更有效地处理预测误差（高信息密度），增强其注视时间效应。补偿机制则体现在中文特有的视觉特征（如长词整体轮廓）能部分补偿语言预测的不足或降低其负荷，例如通过促进副中央凹整体识别来提高跳读概率（词长效应反转）。此外，贝叶斯模型结果表明，认知系统会根据任务需求动态调整对不同通道信息的依赖权重：在跳读决策中，语言预测通道提供的'认知成本'信号（信息密度）被赋予极高权重，强烈抑制跳读；而在决定注视时长时，则需综合评估来自两个通道信息（信息

密度、可预测性、笔画数、词长）构成的整体认知负荷。因此，中文阅读中的认知资源分配是语言预测与视觉编码双通道在资源有限性约束下，通过竞争资源、协同增效、补偿不足、动态加权等方式共同作用的结果。这一模型不仅为深入解析汉语文本认知接收机制（如词长悖论、笔画数效应与信息密度的复杂交互）提供统一框架，揭示中文书写系统特性（表意文字、无空格、高视觉复杂度）对认知加工的根本塑造作用，也为构建普适性跨语言阅读理论（如E-Z Reader模型的汉化）提供关键机制解释与参数依据（如引入文字系统参数、双通道交互规则），推动认知模型的"参数化适应"范式。

本研究提出的认知资源分配框架目前聚焦于视觉文本加工，未来需拓展至多模态场景（如结合听觉ERP、视听双任务范式），探索信息密度的跨通道整合机制；同时，可采用神经成像技术（如EEG/fMRI）实时考察信息密度与神经活动的关联，深入揭示中文阅读中预测与验证的神经机制。在方法学上，可评估新一代大语言模型（如GPT-4.5, ERNIE, RoBERTa）的惊异度计算效果 (Gauthier and Levy, 2019)，进行横向比较以提升评估稳健性；并超越词汇层面的静态分析，结合眼动数据的时序属性（采用EMD分解、Granger因果检验或DTW等方法），探索句子级别惊异度随时间变化的动态特征（如峰值、过渡与前瞻机制），从而更精准刻画认知时间轨迹。值得注意的是，视觉编码路径（体现为词长、笔画数的独立效应及其与信息密度的显著交互）在解释认知负荷方面具有语言预测不可替代的作用，是对语言预测路径的必要结构性补充。然而，为构建更全面且具有认知完备性的"最简充分"模型，未来研究需进一步纳入其他文本结构变量（如依存距离、语法范畴、信息焦点位置等），检验它们与惊异度的交互作用，验证"视觉+语言"双通道模型的解释力及其向其他语言系统迁移的潜力。

参考文献

Marc Brysbaert, Denis Drieghe, and F. Vitu. 2005. Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53-77). Oxford University Press.

V. Clara, B. Corominas-Murtra, and R. Solé. 2021. Matching information demands in language with human cognitive constraints. *Current Opinion in Behavioral Sciences*, 38, 69-76.

C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, and S. Lautenbach. 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46.

D. Drieghe, K. Rayner, and A. Pollatsek. 2005. Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 954-969.

R. Engbert, A. Nuthmann, E. M. Richter, and R. Kliegl. 2005. SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777-813.

K. Friston. 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138.

S. L. Frank and T. F. Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

S. L. Frank, L. J. Otten, G. Galli, and G. Vigliocco. 2015. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140, 1-11.

R. Futrell, E. Gibson, H. Tily, and I. A. Blank. 2021. Predictability and syntactic structure in naturalistic sentence comprehension. *Cognition*, 213, 104688.

J. Gauthier and R. Levy. 2019. Linking artificial and human neural representations of language. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 529-539).

E. Gibbon. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, Language, Brain* (pp. 95-126).

第二十四届中国计算语言学大会论文集，第537页–第550页，济南，中国，2025年8月11日至14日。

D. Genzel and E. Charniak. 2023. A cross-linguistic pressure for uniform information density in word order. *Transactions of the Association for Computational Linguistics*, 11, 528–546.

J. Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics* (pp. 159-166).

E. Huber, S. Sauppe, A. Isasi-Isasmendi, I. Bornkessel-Schlesewsky, P. Merlo, and B. Bickel. 2024. Surprisal From Language Models Can Predict ERPs in Processing Predicate-Argument Structures Only if Enriched by an Agent Preference Principle. *Neurobiology of Language*, 5(1), 167–200.

A. W. Inhoff and K. Rayner. 1986. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40(6), 431-439.

N. Jiang, F. Y. Hou, and X. Jiang. 2020. Analytic versus holistic recognition of Chinese words among L2 learners. *The Modern Language Journal*, 104(3), 567580.

F. Ke, R. Liu, Z. Sokolikj, I. Dahlstrom-Hakki, and M. Israel. 2024. Using eye-tracking in education: review of empirical research and technology. *Educational Technology Research & Development*, 72, 1383–1418.

R. Kliegl, E. Grabner, M. Rolfs, and R. Engbert. 2004. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16(1-2), 262-284.

R. Kliegl, A. Nuthmann, and R. Engbert. 2006. Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1), 12-35.

R. Levy. 2008. Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.

R. Levy and T. F. Jaeger. 2007. Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems* (pp. 849-856).

X. Li, K. Bicknell, P. Liu, W. Wei, and K. Rayner. 2014. Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General*, 143(2), 895-913.

X. Li, P. Liu, and K. Rayner. 2011. Eye movement guidance in Chinese reading: Is there a preferred viewing location? *Vision Research*, 51(10), 1146-1156.

X. Li, K. Rayner, and K. R. Cave. 2009. On the segmentation of Chinese words during reading. *Cognitive Psychology*, 58(4), 525-552.

F. Liang, Y. Liu, F. He, L. Feng, Z. Wang, and X. Bai. 2024. Visual complexity effect in Chinese incidental word learning: Evidence from number of strokes and word length. *Acta Psychologica Sinica*, 56(12), 1734-1750.

S. P. Liversedge, D. Drieghe, X. Li, G. Yan, X. Bai, and J. Hyönä. 2016. Universality in eye movements and reading: A trilingual investigation. *Cognition*, 147, 1-20.

G. Ma and X. Li. 2015. How character complexity modulates eye movement control in Chinese reading. *Reading and Writing*, 28(6), 747-761.

C. Meister, T. Pimentel, and R. Levy. 2021. Revisiting the uniform information density hypothesis. arXiv preprint, arXiv:2109.11635.

R. M. O'Brien. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673–690.

J. Pan, M. Yan, E. M. Richter, H. Shu, and R. Kliegl. 2022. The Beijing Sentence Corpus: A Chinese sentence corpus with eye movement data and predictability norms. *Behavioral Research*, 54, 1989-2000.

C. A. Perfetti and L. N. Harris. 2013. Universal reading processes are modulated by language and writing system. *Language Learning and Development*, 9(4), 296-316.

C. A. Perfetti and L. H. Tan. 1999. The constituency model of Chinese word identification. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 115-134). Lawrence Erlbaum Associates.

K. Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.

K. Rayner. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506.

K. Rayner, T. Warren, B. J. Juhasz, and S. P. Liversedge. 2004. The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1290-1301.

E. D. Reichle, S. P. Liversedge, D. Drieghe, H. I. Blythe, H. S. Joseph, S. J. White, and K. Rayner. 2013. Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, 33(2), 110-149.

E. D. Reichle, T. Warren, and K. McConnell. 2009. Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16(1), 1-21.

C. E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

N. J. Smith and R. Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319.

A. Staub. 2015. The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, 9(8), 311-327.

A. Staub and A. Benatar. 2013. Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin & Review*, 20(6), 1304-1311.

A. Staub and K. Rayner. 2007. Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327–342). Oxford University Press.

B. G. Tabachnick and L. S. Fidell. 2013. Using multivariate statistics (6th ed.). Pearson.

W. Wei, X. Li, and A. Pollatsek. 2018. Word properties in Chinese reading: Some findings from studies of eye movements. In I. Takashima & J. Takahashi (Eds.), *Reading: Assessment, Comprehension and Teaching* (pp. 121-156). Nova Science Publishers.

V. Whitford and M. F. Joanisse. 2021. Eye movement control in reading: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(9), 1432-1449.

E. G. Wilcox, J. Gauthier, J. Hu, P. Qian, and R. Levy. 2020. On the predictive power of neural language models for human real-time comprehension behavior. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1707-1713).

G. Yan, H. Tian, X. Bai, and K. Rayner. 2010. The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, 101(1), 21-36.

M. Yan and R. Kliegl. 2021. The Beijing Sentence Corpus. https://doi.org/10.17605/OSF.IO/VR3K8

M. Yan, R. Kliegl, E. M. Richter, A. Nuthmann, and H. Shu. 2010. Flexible saccade-target selection in Chinese reading. *Quarterly Journal of Experimental Psychology*, 63(4), 705-725.

M. Yan, W. Zhou, H. Shu, and R. Kliegl. 2014. Perceptual span depends on font size during the reading of Chinese sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1677-1686.

H. M. Yang, G. W. McConkie, and Y. F. Yang. 2009. The impact of visual factors on orthographically complex characters in reading Chinese. *Writing Systems Research*, 1(2), 127-137.

C. Zang, S. P. Liversedge, X. Bai, and G. Yan. 2011. Eye movements during Chinese reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 961-978).

第二十四届中国计算语言学大会论文集，第537页-第550页，济南，中国，2025年8月11日至14日。