

线性混合效应模型在二语眼动阅读研究 数据分析中的应用

——以 R 语言为例

易 维 鹿士义

北京大学

提 要 近年来,眼动技术在应用语言学和二语习得领域的应用越来越广泛。眼动技术有利于揭示阅读理解和语言加工过程中的认知心理机制。然而,这对研究者的数据分析能力也提出了新的挑战。随着研究的转型,传统的方差分析方法逐渐被线性混合效应模型(linear mixed-effects model)所取代,而使用 R 语言这样的开源语言进行数据分析在研究者中日渐流行。本文以 R 语言为统计软件,详细介绍了使用线性混合效应模型分析眼动数据的原理、步骤和方法,同时也为使用该模型分析其他类型的数据提供了一定的借鉴。

关键词 混合效应模型 眼动数据 R 语言 统计分析

一 眼动技术和眼动指标

阅读过程中的基本眼动行为包含眼跳(saccade)和注视(fixation)。眼跳是眼睛从已知区域向未知区域运动,而注视则是眼睛保持相对静止的状态,即两次眼跳之间的停顿。一般认为,阅读过程中,人们主要是从注视中获取信息的(闫国利等 2013)。阅读拼音文字时,平均眼跳距离为 7 到 9 个字母,平均注视时间在 200 到 300 毫秒左右。相比之下,阅读中文文本时,平均眼跳距离为 2 到 3 个汉字,平均注视时长与阅读拼音文字相比差别不大(Zang et al. 2011)。眼睛除了做前进式的运动,有时也会退回到之前阅读过的部分,即“回视”。熟练阅读者的回视一般占到 10%~15%左右。除了回视,阅读过程中还存在大量的跳读现象。例如,在英语里,大约 15%的实词可能被跳读。相比之下,高达 65%的虚词可能被读者跳过(Just & Carpenter 1980)。研究表明,阅读中的眼动行为受到形、音、义等多方面因素的影响(Juhasz & Pollatsek 2011; Rayner & Liversedge

2011),且在词语加工、语块加工(Yi et al. 2017)和句子加工层面都有所反映。在中文阅读领域,大量研究表明,中文阅读受到词边界、汉字/词形复杂度、汉字/词语频次、词语可预测性等因素的影响(Zang et al. 2011)。

眼动研究借助眼动仪来收集数据。被试在眼动仪前完成实验任务后,研究人员通常要根据研究目标确定一个或多个目标区域,即“兴趣区”(interest area),从而获取眼动仪在该区域收集到的数据,以便进行统计分析。阅读研究中,兴趣区一般是以词为单位,有时也可以是更大的区域(如短语/语块或句子)。在视知觉研究或视觉情景范式(visual-world paradigm)中,兴趣区则往往是图片或者图片中的特定区域。眼动数据包括空间和时间两个维度(闫国利等 2013)——前者涉及眼跳或者注视的位置、距离和次数,后者反映的则主要是不同阶段(如首次注视、凝视/第一遍阅读、第二遍阅读)的注视时长。眼动仪收集到的数据表现为不同的眼动指标(eye-tracking measures)。根据时间进程的不同,眼动指标又可以分为早期指标和晚期指标。早期的眼动指标通常包括首次注视时间(first fixation duration)、凝视时间/第一遍阅读时间(gaze duration/first-pass reading time)和跳读率(skipping rates),反映的主要是语言加工的早期阶段(如字形识别、词汇通达)。晚期的眼动指标包括第二遍阅读时间(second-pass reading time)、回视路径时间(regression path duration/go-past time)、总阅读时间(total reading time/total fixation duration)、注视次数(fixation count)等,反映的主要是语言加工过程中的后期阶段,例如信息再认、信息整合等(Juhasz & Pollatsek 2011; Roberts & Siyanova-Chanturia 2013)。尽管不同的眼动指标可以反映不同的语言加工步骤,需要指出的是,眼动指标与特定的认知操作之间并不存在简单的一一对应关系(Rayner & Liversedge 2011)。在选择眼动指标时,至少有以下三点值得研究者注意。第一,建议选择多个眼动指标。Rayner(1998)曾指出,仅仅使用单一指标很难反映出认知加工的实际情况。多个眼动指标能够从不同方面提供多方证据,有利于研究者细致剖析认知加工的过程。第二,选择眼动指标时要有相应的理据。具体而言,研究者决定纳入的眼动指标要能支持其研究目标,并提供相应的文献依据(Godfroid & Hui 2020)。第三,不同类型的兴趣区所适用的眼动指标也可能有所不同。例如,当兴趣区是词语时,首次注视时间是一个很好的指标。但是,当兴趣区是大于词的单位时(如短语/语块),这个指标的作用就非常有限。此时,研究者最好能同时纳入凝视时间和第二遍阅读时间这两个指标(Rayner 1998)。常用的眼动指标及其定义详见表 1。

眼动技术建立在“眼-脑假说”基础之上(Just & Carpenter 1980),即眼动反映的是大脑实时加工的情况。具体而言,当前被注视的项目就是大脑正在加工的内容,而对该项目的注视时长反映的则是加工该项目所需的认知努力。近年来,眼动技术越来越广泛地被应用于应用语言学和二语习得研究领域。较之于其他实验技术(如自定速阅读),眼动

表1 常用眼动指标及其定义

加工阶段	眼动指标	定义
早期	首次注视时间 first fixation duration	兴趣区内第一个注视点的时长
	凝视时间/第一遍阅读时间 gaze duration/first-pass reading time	兴趣区在第一遍阅读中所有注视点时长的总和
	跳读率 skipping rates	兴趣区在第一遍阅读中被跳读的可能性
晚期	第二遍阅读时间 second-pass reading time	离开兴趣区后,再次回到兴趣区时所有注视点的时长总和
	回视路径时间 regression path duration/go-past time	离开兴趣区阅读兴趣区后面的内容之前,兴趣区内所有注视点时长的总和(包括回视时间在內)
	总阅读时间 total reading time/total fixation duration	兴趣区内所有注视点的时长总和
	注视次数 fixation count	兴趣区被注视的总次数

注:凝视时间和第一遍阅读时间这两个术语基本含义相同,针对的分别是兴趣区为词和短语的情况。

技术用于语言研究时具有诸多优势。首先,眼动仪能够在不干扰自然阅读的情况下收集数据,因而能保证阅读任务的自然性(生态性)。其次,眼动仪能实时收集到非常丰富眼动数据。通过选择不同的眼动指标,研究者能够对语言使用者的在线语言加工进行深入的分析。尤其值得指出的是,在某些研究设计中,如果配合使用早期和晚期眼动指标,研究者还能一定程度上界定/区分认知加工的时间进程。最后,眼动技术支持多模态研究。除了文本阅读外,研究者还能使用眼动技术研究听觉语言加工、视频/图片浏览以及真实/虚拟场景中的互动行为。

二 线性混合效应模型

统计分析旨在根据实验获取的样本数据进行建模,从而帮助研究者对总体背后的规律进行推断。所谓建模,本质上就是选择合适的数学模型对离散分布的数据进行拟合。合乎规范的统计分析,其结果要能解释数据中的离散变化。应用语言学和二语习得领域的统计分析,大部分情况下只涉及线性模型(如方差分析、多元回归)。眼动研究通常是实证性的,即操作若干个自变量(实验条件),观察因变量(实验结果),并通过统计分析探究自变量是否对因变量的变化具有解释力。多年来,方差分析作为一种统计手段一直被

广泛应用于实验心理学等领域。方差分析本质上属于多元回归模型,即使用多个自变量来解释因变量。使用方差分析时,研究者需要按被试或实验材料对原始数据取均值,并在这些均值的基础上建模分析。因此,在报告统计结果时,方差分析往往分为基于被试的分析(F_1 检验)和基于材料的分析(F_2 检验)。尽管方差分析的应用非常广泛,它的局限性近年来也逐渐引起人们的注意。首先,方差分析针对的仅仅是因变量的均值,因而无法解释因变量中的个体差异。其次,方差分析只适用于自变量为分类变量(如性别)的情形。当自变量中包含连续变量(如身高)时,方差分析就不适合了。再次,实验数据(因变量)往往存在基于被试和材料的随机变化,然而方差分析无法将这两种随机效应同时考量进去。研究者往往只是在 F_1 或者 F_2 检验中——而非同时在 F_1 和 F_2 检验中——发现某个自变量的显著效应,这就给统计结果的解读带来极大的问题。

不同于方差分析,线性混合效应模型在分析数据时不需要对原始数据进行均值处理,也允许出现数据缺失。同时,线性混合效应模型也能纳入不同类型的自变量和因变量。线性混合效应模型,顾名思义,是将数据(因变量)中的离散变化看成“固定效应”和“随机效应”的混合,由固定效应和随机效应共同解释因变量的变化。固定效应和随机效应的区别,主要取决于某些变量在抽样过程中是否被穷尽。比如说,研究者招募了30名汉语二语者被试,并让他们对20个双字组合进行词汇判断,即尽可能快速地通过按键反应判断这些字符是不是词。被试中一半是男性,一半是女性。研究者感兴趣的问题是,在控制其他变量的情况下,汉语二语者的词汇判断反应时是否存在性别差异。如果要使用混合效应模型对这个实验收集到的数据进行分析,那么,“性别”属于固定效应,而被试和材料(双字组合)属于随机效应。性别这个变量只有男性和女性这两类,而被试和材料仅仅是从总体(所有的二语者/所有的中文双字组合)中抽样出来的。随机效应又能够进一步分为随机截距(radom intercept)和随机斜率(radom slope)。随机截距表示的是平均值的随机变化,而随机斜率表示的是效应量的随机变化。以上面的假想实验为例,如果研究者确实发现了词汇判断反应时存在显著的性别差异,那么,除了揭示出这个总体趋势,通过混合效应模型,我们还能具体到每一个被试,每一个实验材料(双字组合),得知性别效应在被试和材料之间的随机变化情况。其中,被试的随机截距代表的是所有被试对双字组合进行词汇判断时的平均用时在不同被试之间的变化,而材料的随机截距表示的是双字组合在所有被试中的平均反应时在不同双字组合之间的变化。相应地,材料的随机斜率代表的是性别效应(男性和女性的词汇判断反应时差异)在实验材料之间的变化(男性和女性的反应时差异大小可能随不同的双字组合而有所变化)。混合效应模型包括线性混合效应模型(linear mixed-effects model)和广义线性混合效应模型(generalized linear mixed-effects model, GLMM)。前者适用于因变量符合正态分布

(normality)的情形,因此使用最为广泛,后者适用于因变量不符合正态分布的情况。当因变量为分类变量时(如被试在某道题上是否答对,或者某个词是否在阅读中被跳读),则可以使用混合效应逻辑斯蒂模型(mixed-effects logistic model);当因变量为个数(如注视次数)时,则可以使用混合效应泊松模型(mixed-effects Poisson model)。限于文章篇幅,本文仅介绍线性混合效应模型在眼动数据分析中的应用。

三 R 语言

统计分析用到的软件种类繁多,比较常用、功能全面(包含数据导入、分析、可视化/绘图,结果报告)的包括 SPSS、R 语言、MATLAB、SAS 和 STATA 等。其中,由于简单易学、开源、免费且有强大的统计分析工具,R 语言近年来成为应用语言学和二语习得领域广泛使用的统计软件。相比于菜单式设计的 SPSS,R 语言能够在统计分析数据的过程中赋予研究者更大的自由。一方面,研究者在选择何种统计方法或绘图手段时有更多的选择,另一方面,哪怕是同样一种统计方法或者绘图手段,研究者也能找到多个不同的统计包(package)来实现。从这个意义上来说,R 语言增强了统计分析过程的透明度和灵活性,同时也让研究者能够通过简单易懂的代码语言,掌握数据分析的自主权。

四 统计分析的流程

统计分析的流程一般分为数据准备、数据探索(exploratory data analysis, EDA)、统计建模、结果解读与报告这四个基本步骤。眼动仪收集的数据非常丰富。为了确保眼动数据的可靠性,研究者在进行眼动实验时需要进行校准,以便让眼动仪能够准确地识别和追踪被试的眼动。实验结束后,研究者还需要对每个被试的眼动数据进行检查,并根据一定的标准筛除不合格被试(例如不认真参与实验、眼动未能捕捉到)或不合格试次(例如该试次的眼动数据丢失)的数据。在此基础上,研究者往往会根据研究目的划分出一个或多个兴趣区,并选择多个眼动指标,使用眼动仪配套的软件导出实验数据。实验数据通常包含注视时间、次数、跳读率等,研究者还可以结合本领域的通常做法,将过短或过长的注视时间剔除。

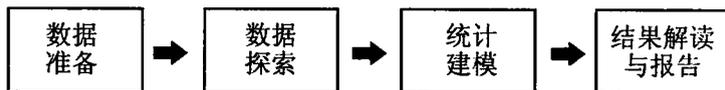


图1 统计分析的一般流程

通常来说,数据准备阶段要完成以下步骤。以线性混合效应模型为例,分析数据之前,首先需要将所有数据合并为一个数据表,并按照逐人、逐个材料整理为长格式(long format)。接下来,研究者需要检查数据中是否有缺失值(missing values)。通常情况下,数据缺失在所难免——有时是因为某些被试的部分实验任务没有完成,有时则是因为部分数据经过筛除之后缺失了。线性混合效应模型允许出现数据缺失,必要的时候,研究者也可以对缺失值进行插补(imputation)。完成了上述步骤之后,研究者需要对各变量进行转换和编码。因为反应时不符合正态分布,所以,通常情况下反应时数据可以考虑进行对数转换(log transformation)。对于分类变量而言,需要确保它们没有被编码成连续变量。统计模型中的截距(intercept)反映的是所有自变量取值为0时因变量的取值。有的自变量没有0这个取值(如年龄、身高),为了使统计结果更容易解读,一般建议对连续变量进行中心化(centering)处理,即将自变量的每个值减去该变量的均值。这样,统计模型的截距就可以解释为当所有自变量取平均值时,因变量的取值。有时候,不同自变量的取值范围相差很大,为了便于进行统计分析,还可以对所有自变量进行标准化(standardization)处理,即将所有取值转换为z分数。这样做还有一个好处,就是研究者可以直观地比较不同自变量对同一个因变量影响的大小。

统计分析是为了揭示自变量和因变量之间的关系,在正式建模之前,一般建议结合绘图手段进行数据探索。数据探索的目标包括:(1)考察各自变量之间的相关性(correlation),尤其是各自变量之间是否存在高度的相关性;(2)了解各自变量与因变量之间的关系,初步明确自变量是否对因变量具有解释力;(3)检查统计假设是否违反。许多研究者在进行统计分析时往往忽视统计假设,这是不可取的。任何一项统计检验都是建立在相应的统计假设基础之上的,如果统计分析违反了重要的统计假设,那么得出的统计结果是不可信的。对线性混合效应模型来说,需要检验的统计假设包括自变量和因变量之间是否符合线性关系(linearity),残差(residuals)是否符合正态分布,以及方差齐性检验(homogeneity of variance)。各项假设检验的具体操作详见下文。

在统计建模阶段,最重要的是要根据因变量的分布性质选择合适的模型进行建模分析。就线性混合效应模型而言,统计建模的方法分为两派,一派是数据驱动型(data-driven)的,一派是研究驱动型(design-driven)的。前者在统计建模时遵循由简入繁(forward selection)或者由繁入简(backward selection)的步骤,通过比较层层嵌套的统计模型(nested models),选择出解释力更强、结构更为经济的模型,并报告最优模型的结果为统计分析的最终结果。后者(Barr et al. 2013)则认为,统计建模应该建立在具体的实验设计基础上,自变量和随机效应的去留并不能取决于模型的比较。由于数据驱动型选出的模型更为经济,本文将在下文用这种方法展示使用混合效应模型分析眼动数据的

具体步骤。最优模型筛选出来之后,研究者需要报告最优模型中的统计结果。如果统计结果中没有提供 p 值,那么研究者还需要通过其他方法计算出 p 值。最后,统计结果一般按照《美国心理学会发表指南(第七版)》(American Psychological Association 2019)的格式进行呈现。除了报告是否存在显著效应,近年来,学界(Norris et al. 2015)还强烈建议研究者计算并报告效应量(effect size)的大小。对重点或者较复杂的统计结果,最好配合统计图表加以说明。

五 使用线性混合效应模型分析二语阅读眼动数据

为了更好地帮助读者理解使用混合效应模型分析眼动数据的方法,在这里,我们以 Yi et al. (under review) 的研究为例,以 R 语言(3.6.2 版本,R Core Team 2015)为统计软件,呈现具体的操作步骤。在这项研究中,研究者考察了词长(word length)、具体性(concreteness)和语境支持度(contextual support)这三个因素对阅读过程中的注意分配和伴随性词汇习得效果的影响。具体来讲,这项研究聚焦于二语者阅读文本时首次接触到二语生词的情形。研究者采用重复测量实验设计,以中高级水平的中国英语学习者为被试,选择了 32 个被试不认识的单语素英语名词作为实验材料/目标词。在这 32 个名词中,一半为长词,一半为短词;一半为具体词,一半为抽象词。每个词被嵌入两个 that 引导的定语从句中,其中一个为高语境(informative),另一个为低语境(neutral)。为了不让被试看到一个词同时出现在高语境和低语境句子中,实验句经过平衡设计(counterbalancing)之后被分成两组,每个被试只阅读其中的一组材料。实验开始前,被试在网上完成一项问卷调查,报告自己的二语学习背景。实验开始后,被试按自己的习惯阅读每个句子,理解句意,并回答阅读理解问题。阅读任务结束后,被试需要紧接着进行若干词汇知识测验(词形识别、词义描述、词义识别)。实验末尾,被试还需要通过一个四点李克特量表(Likert scale)报告自己在参加实验前对每个目标词的熟悉程度(1:我从未见过这个词;2:我见过这个词;3:我认识这个词;4:我对这个词非常熟)。研究者感兴趣的问题包括:(1)词长、具体性和语境支持度对二语者阅读过程中的生词在线加工有何影响?(2)词长、具体性和语境支持度对二语者首次接触目标生词后所获得的词汇知识有何影响?(3)二语者对生词的注意加工是否可以预测他们的伴随性词汇习得成效?由于篇幅所限,本文仅介绍 Yi et al. 一文中探讨生词在线加工的部分。

5.1 数据准备

研究者将目标词和目标词后面的定语从句划分为两个不同的兴趣区,并选择了早期眼动指标和晚期眼动指标。其中,早期指标包括首次注视时间、凝视时间和跳读率,晚期指标则包括第二遍阅读时间、总阅读时间和注视次数。按照逐人、逐词的方式,眼动数据

(包含被试信息、目标词信息和各项眼动指标)被整理成为长格式(图 2)。其中,两位被试的阅读理解题正确率低于 75%,这表明他们可能存在阅读理解方面的障碍或者没有认真参与实验、理解句意。据此,研究者将这两位被试的数据整体剔除。此外,有 2.3%的试次没有捕获到眼动数据,因此这些试次也没有纳入统计分析中来。参考前人(Betancort et al. 2009)的研究,注视时长低于 80 毫秒的数据也进行了删除(7.4%)。由于被试只是在阅读理解的过程中接触到目标生词,且生词没有复现,绝大部分被试无法完成词义描述任务,因此该任务也不纳入统计分析。值得注意的是,该研究使用的是真词,而非假词,因此每个被试在实验结束后报告的对每个目标词的事前熟悉度(prior exposure)被用来进一步剔除词义测验数据。具体来说,如果某个被试报告他/她在实验开始之前就已经认识某个目标词(即李克特量表回答 ≥ 3),且他/她在所有词汇知识测验中确实都答对了,那么该被试在这个生词上的所有眼动数据和词汇测验数据都不会纳入统计分析。

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	被试编号	材料编号	词长	具体性	语境支持度	平均双字符频次	事前熟悉度	首次注视时间	凝视时间	跳读率	第二遍阅读时间	总阅读时间	注视次数
2	35	4	9	621	1	1.76	0	176	1005	0	308	1313	5
3	8	16	7	593	0.8	2.69	0	82	82	0	762	980	5
4	35	25	6	242	0.1	2.1	0	493	618	0	466	1084	4
5	35	26	6	379	0.4	2.61	0	603	603	0	644	1247	5
6	35	27	6	367	0.2	2.67	0	217	385	0	775	1763	7
7	35	11	6	577	0.6	2.71	0	215	878	0	203	1081	5
8	35	20	8	282	0.1	2.59	0	104	104	0	882	1944	9
9	8	30	7	376	0	2.74	0	374	374	0	376	750	3
10	65	11	6	577	0.6	2.71	1	426	426	0	330	1310	4
11	35	17	8	242	0	2.35	0	852	1128	0	204	1808	6
12	35	8	8	573	0.9	1.63	0	363	860	0	NA	860	3
13	65	18	9	324	0.3	1.77	1	287	408	0	55	1662	8
14	65	27	6	367	0.2	2.67	1	207	673	1	207	1320	5
15	65	26	6	379	0.4	2.61	0	182	182	0	574	1374	5
16	35	19	8	283	0.3	2.18	0	254	408	0	742	1902	8
17	35	10	7	487	0.8	1.9	0	248	740	0	655	1855	8
18	65	8	8	573	0.9	1.63	0	118	118	0	822	2053	7
19	35	1	10	604	0.8	2.22	0	230	1035	0	441	2196	10
20	65	16	7	593	0.8	2.69	0	501	501	1	233	1831	7

图 2 长格式数据

由于注视时间不符合正态分布,所以研究者对注视时间(首次注视时间、凝视时间、第二遍阅读时间、总阅读时间)取自然对数,进行了转换。词长(字母数)和语境支持度(基于被试评分,百分比为单位)这两个自变量被编码为连续变量。具体性为分类变量(具体词、抽象词),采用的是虚拟编码(dummy coding),以抽象词为参照组。被试在实验结束后报告的对每个目标生词的事前熟悉度也被作为一个协变量(covariate),分为“见过”(李克特量表回答=2)和“没见过”(李克特量表回答=1)两个水平,并以“没见过”为参照水平,进行虚拟编码。除此以外,由于实验材料没有完全控制平均双字符频次(mean bigram frequency),且平均双字符频次可能影响词语——尤其是低频词的加工,所以平均双字符频次作为一个协变量也纳入了统计分析。所有的数值变量(numeric variables)都进行了中心化处理。

5.2 数据探索

研究者使用 R 语言对各个自变量之间的相关性进行了分析,结果(图 3)显示,词长、

具体性和语境支持度之间的相关性很弱(相关系数小于 0.1)。平均双字符频次与词长之间存在中度负相关($r = -.59$)关系。以首次注视时间为例,绘图(图 4)显示,使用眼动数据进行建模时,符合线性关系假设,且残差分布均匀,没有违反方差齐性(图 5)。同时,QQ 图(图 6)显示,残差基本上符合正态分布。

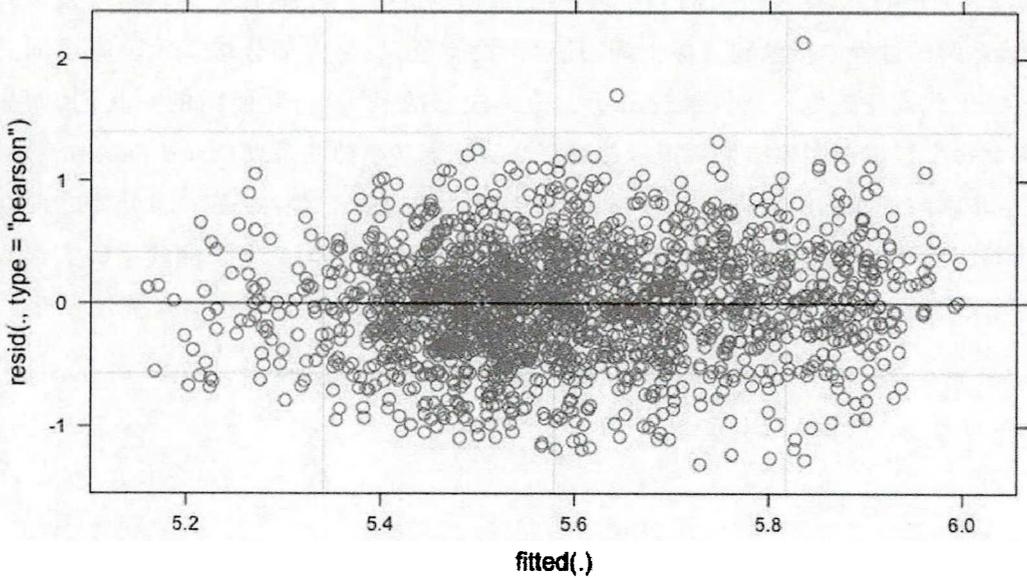


图 3 自变量之间的相关性

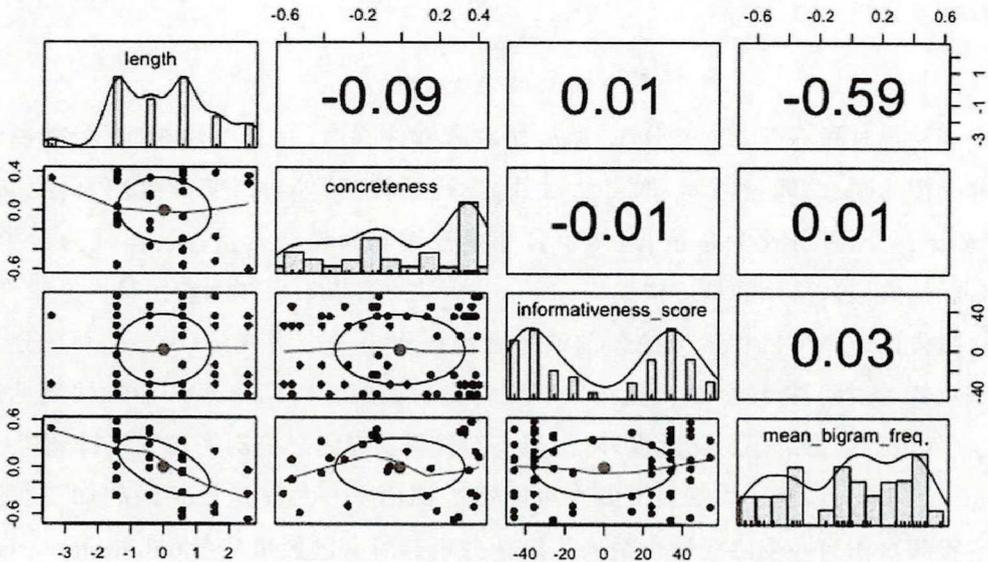


图 4 检验线性假设

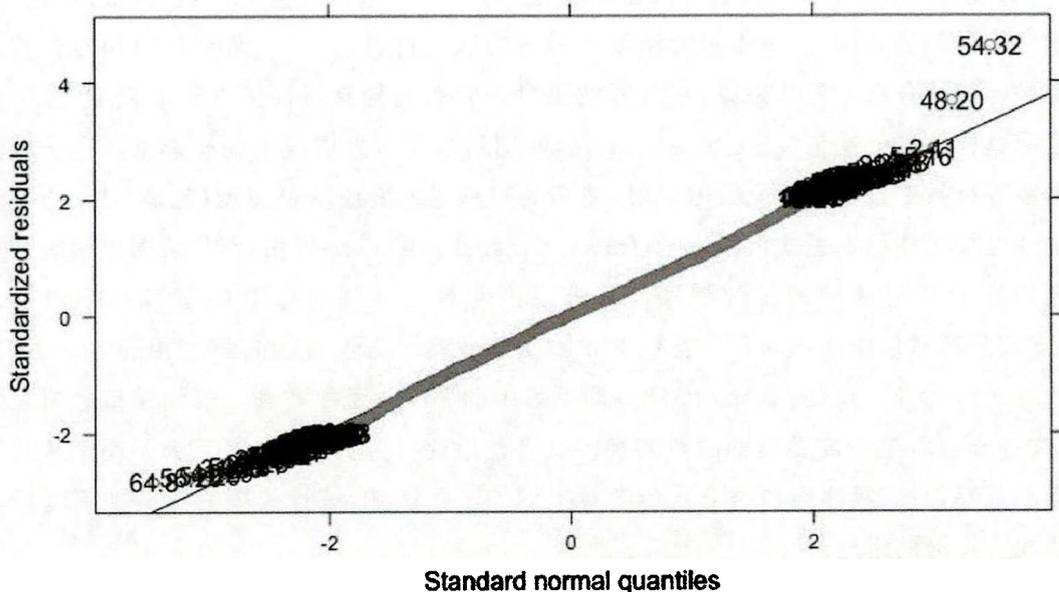


图5 检验方差齐性假设

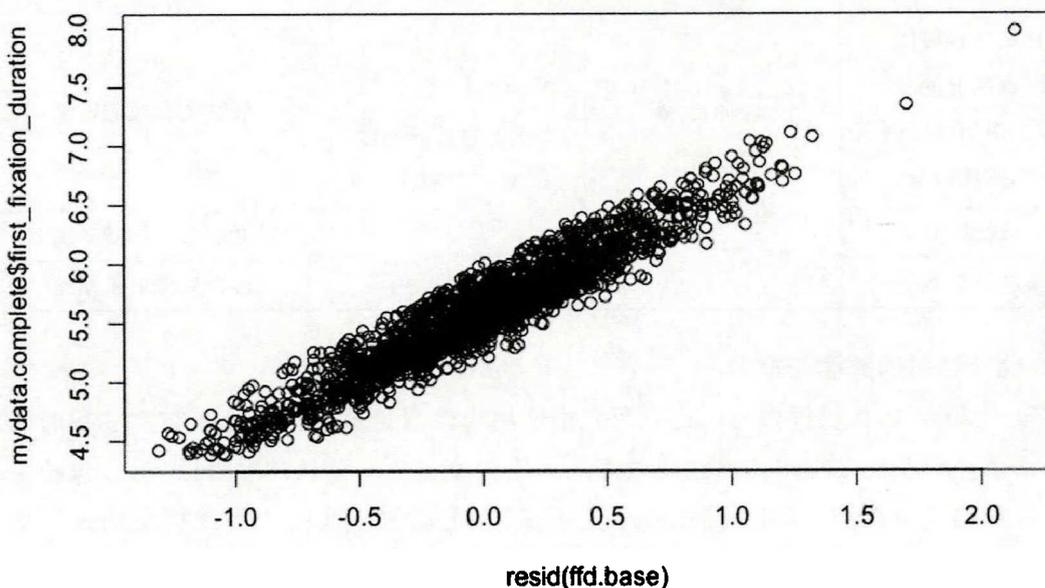


图6 检验正态分布假设

5.3 统计建模

为了回答词长、具体性和语境支持度对阅读过程中生词首次加工的影响,研究者使用线性混合效应模型进行了统计分析。其中,词长、具体性和语境支持度为三个自变量,早期加工眼动指标(首次注视时间、凝视时间、跳读率)和晚期加工眼动指标(第二遍阅读时间、总阅读时间、注视次数)为因变量。协变量包括生词的事前熟悉度和双字符频次。

为了提升统计检验力(power),词长和语境支持度处理为连续变量,而具体性处理为分类变量(具体词、抽象词)。由于实验涉及不同的被试和材料(生词),研究者使用混合效应模型进行统计分析。具体来说,在分析注视时间时选用的是线性混合效应模型,在分析跳读率这样的二元变量(binary variable)时使用的是混合效应逻辑斯蒂模型。在分析注视次数等计数变量(count variable)时,使用的则是混合效应泊松模型(见表2)。研究者在统计建模时使用递进式模型筛选(forward-model selection)方法,即从最简单的统计模型逐步递进到更复杂的统计模型。统计建模时首先从随机截距模型(intercept-only model)起步,然后依次放入每个自变量和协变量。模型比较通过R语言中的lme4程序包(1.1-21版本, Bates et al. 2015),利用其中的anova函数实现。最优模型选出来之后,进一步查看是否需要增加随机斜率。显著性检验设置为.05水平。为了便于从时间进程上了解词长、具体性和语境支持度对生词加工的影响,统计建模分早期加工指标和晚期加工指标两组进行。

表2 常用眼动指标适用的混合效应模型

因变量	自变量	协变量	统计模型
首次注视时间 凝视时间 第二遍阅读时间 总阅读时间	词长 具体性 语境支持度	生词的事前熟悉度 双字符频次	线性混合效应模型
跳读率			混合效应逻辑斯蒂模型
注视次数			混合效应泊松模型

5.4 统计结果的解读与报告

限于篇幅,本文仅报告 Yi et al. 研究中使用线性混合效应模型分析首次注视时间的统计结果。模型比较得出首次注视时间的最优模型包含了词长的显著效应,并包括了被试和材料的随机截距。使用 summary 这个函数,可以快速地读取最优模型的统计结果(图7)。

具体来说,各部分统计结果的解读如下:

Formula:最佳模型的公式。这个公式显示了最佳模型中有显著效应的变量以及随机效应。以首次注视时间为例,这个最佳模型包含了词长的主效应、被试的随机截距和材料的随机截距。最佳模型的公式建议报告出来。

```

> summary(ffd.best_model)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: first_fixation_duration ~ length + (1 | subject) + (1 | item)
Data: mydata.complete

      AIC      BIC  logLik deviance df.resid
2487.9  2515.4 -1238.9  2477.9    1815

Scaled residuals:
    Min      1Q  Median      3Q      Max
-2.8406 -0.6414  0.0213  0.6451  4.6186

Random effects:
 Groups   Name      Variance Std.Dev.
subject  (Intercept) 0.029420 0.17152
item     (Intercept) 0.001262 0.03553
Residual                    0.215125 0.46382
Number of obs: 1820, groups: subject, 63; item, 32

Fixed effects:
              Estimate Std. Error t value
(Intercept)  5.587476   0.025026 223.267
length       -0.025477   0.008898  -2.863

Correlation of Fixed Effects:
      (Intr)
length -0.001

```

图7 最优模型的统计结果

AIC、BIC、logLik、deviance、df. resid:这部分报告的是最佳模型的拟合情况,可以作为衡量模型拟合度的参考,不需要报告。

Scaled residuals:这里报告的是残差,也就是最佳模型的预测值与数据表中实际值之间的差距,不需要报告。

Random effects:这里报告的是最佳模型的随机效应。具体来说,包含了基于被试的随机截距和基于材料的随机截距,也就是最佳模型预测的平均首次注视时间在所有被试和材料之间的个体差异。

Number of obs: 这里报告的是最佳模型调用的数据量,包括行数(1820)、被试数量(63人)和材料/生词数量(32个)。

Fixed effects:固定效应,也就是最佳模型中包括的显著效应。包括截距——也就是平均首次注视时间和词长的主效应。

Correlation of Fixed Effects: 固定效应之间的相关性,不需要报告。

由于 Yi et al. 的研究对首次注视时间进行了对数转换,因此在解读每个变量的效应时,必须用 $\exp()$ 函数转换回到毫秒单位上来。例如,截距的回归系数为5.587,这就表明当词长取平均值的时候(注意:在数据处理阶段,词长和阅读成绩都做了中心化处理,因

此当它们取值为0的时候,即代表它们的值为各自的平均值),所有被试在所有生词上的平均首次注视时间为267毫秒 $[\exp(5.59)]$ 。词长效应的回归系数为-0.025,这表明词长每增加一个字母,首次注视时间会减少3% $[1-\exp(-0.03)]$ 。随机效应这部分, σ^2 表示的是残差,可以不报告。接下来, τ^{00} subject 和 τ^{00} item 分别表示的是被试和材料的随机效应(截距)的大小。也就是说,平均首次注视时间在被试之间的浮动范围为0.03,在材料(生词)之间的浮动范围接近于0。由于实验设计本身涉及不同的被试和不同的生词,因此,按照某些统计学家(Barr et al. 2013)的看法,不论随机效应的效应量大小,被试和材料的随机效应都应当纳入统计模型中来。统计结果一般通过表格的形式加以呈现,表格参照美国心理学会(APA)的格式要求。在R语言中,通过安装并调用sjPlot这个程序包(Lüdtke 2019)中的tab_model()命令,可以自动绘制符合APA规范的表格,并获得每个效应的p值。根据论文发表的格式要求,可以增删其中的要素或者调整表格的格式。最后,我们可以参考下文来报告统计结果:

研究者利用R语言中的lme4程序包和sjPlot程序包,通过线性混合效应模型(linear mixed-effects model),递进式(forward-model selection)建模,逐步加入词长、具体性、语境支持度等自变量和其他协变量(事前熟悉度、双字符频次)。统计结果表明,词长(Estimate = -0.03, SE = 0.01, t = -2.863, p = .004)对二语者阅读生词时的首次注视时间具有显著影响。具体来说,词长每增加一个字母,二语者的首次注视时间会下降3% $[1-\exp(-0.03)]$ 。

六 结语

本文系统地介绍了使用线性混合效应模型分析眼动数据的原理和方法,在眼动指标的选择和统计流程与规范等方面也给出了明确的意见。虽然本文是从眼动数据分析视角切入的,但文中关于统计分析的一般流程以及线性混合效应模型的介绍也适用于其他类型数据的统计分析。我们希望本文有助于应用语言学和二语习得研究领域的国内同行(尤其是硕士、博士研究生)提升统计素养,提升统计分析的规范性,从而为高水平的学术研究扫除这方面的障碍。

参考文献

- 闫国利、熊建萍、臧传丽等(2013)阅读研究中的主要眼动指标评述,《心理科学进展》第4期,589—605页。

- American Psychological Association (2019) *Publication Manual of the American Psychological Association*. 7th edition. Washington, DC: American Psychological Association.
- Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory & Language*, 68(3), 255–278.
- Batse, D., Mächler, M., Bolker, B. M. & Walker, S. C. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Betancort, M., Carreiras, M. & Sturt, P. (2009) The processing of subject and object relative clauses in Spanish: An eye-tracking study. *The Quarterly Journal of Experimental Psychology*, 62(10), 1915–1929.
- Godfroid, A. & Hui, B. (2020) Five common pitfalls in eye-tracking research. *Second Language Research*, 36(3), 277–305.
- Juhász, B. & Pollatsek, A. (2011) Lexical influences on eye movements in reading. In Liversedge, S. P., Gilchrist, I. D. & Everling, S. (eds.). *The Oxford Handbook of Eye Movements*, 873–894. New York: Oxford University Press.
- Just, M. A. & Carpenter, P. A. (1980) A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354.
- Lüdecke, D. (2019) sjPlot: Data visualization for statistics in social science. R package version 2.7.2. <https://CRAN.R-project.org/package=sjPlot>. (访问日期:2021年11月2日)
- Norris, J. M., Plonsky, L., Ross, S. J. & Schoonen, R. (2015) Guidelines for reporting quantitative methods and results in primary research. *Language Learning*, 65(2), 470–476.
- Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Rayner, K. & Liversedge, S. P. (2011) Linguistic and cognitive influences on eye movements during reading. In Liversedge S. P., Gilchrist, I. D. & Everling S. (eds.). *The Oxford Handbook of Eye Movements*, 751–766. New York: Oxford University Press.
- Roberts, L. & Siyanova-Chanturia, A. (2013) Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, 35, 213–235.
- Yi, W., Lu, S. Y. & DeKeyser, R. Initial processing and learning of novel words during reading: Orthographic, semantic, and contextual influences (under review).
- Yi, W., Lu, S. Y. & Ma, G. J. (2017) Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, 33(4), 519–549.
- Zang, C. L., Liversedge, S. P., Bai, X. J. & Yan, G. L. (2011) Eye movements during Chinese reading. In Liversedge, S. P., Gilchrist, I. D. & Everling, S. (eds.). *The Oxford Handbook of Eye Movements*, 961–978. New York: Oxford University Press.

作者简介

易维,北京大学对外汉语教育学院助理教授,主要研究方向为语言认知与习得。

Email:weiyisla@pku.edu.cn。

鹿士义,北京大学对外汉语教育学院研究员、长聘副教授,主要研究方向为语言认知与习得、语言测试。Email:lushiyi@pku.edu.cn。

WANG, Yuan: The Framework System and Teaching Strategies for the Discrimination of “Possible Words” in Modern Chinese

This paper takes the modern Chinese words with “possible meaning” as the research object, defining “possible words” according to the cognitive motivation of “possible”, and dividing “possible words” into eight categories from the perspective of the semantic category source of “possible meaning”. In view of the characteristics of “possible words” such as great quantity, detailed meaning and close meaning, and in order to distinguish the usages of “possible words”, this paper establishes a discrimination framework system from the following three aspects: semantics, syntax and pragmatics. With the guidance of this discrimination framework theory, the author analyzes the similarities and differences in the usage of “possible words” in the near sense and demonstrates the teaching ideas with some examples.

Key words: possible words, discrimination framework, teaching strategy

YI, Wei & LU, Shiyi: Linear Mixed-effects Model for the Analysis of Eye-tracking Data in L2 Reading Research: A tutorial in R

Recent years have witnessed an increase in the use of eye tracking in applied linguistics and second language acquisition research. With the aid of eye tracking, cognitive mechanisms underlying reading and language processing can be revealed. However, this brings a new challenge to researchers, especially in data analysis. A notable change in statistical analysis practice is that analysis of variance (ANOVA) has been gradually replaced by linear mixed-effects model. Meanwhile, more and more researchers have turned to open-source software, such as R. This tutorial introduces the principles and procedures while analyzing eye-movement data in R using linear mixed-effects model, hoping to set an example for the analysis of other types of data as well.

Key words: linear mixed-effects model, eye-tracking data, R, data analysis

CUI, Yonghua: Dialogue on the Cultivation of Chinese Intercultural Communicative Competence

This is the speech prepared by the author for the webinar “‘Intercultural Chinese Teaching’ Workshop: Cloud Experience and Conversation Series” held by the College of International Education, Shandong University. This speech argues that the goal of teaching Chinese as a second language is to cultivate intercultural communicative