中文词切分的依据及眼动控制模型的解释*

向慧雯[1] 胡欣灼[1] 杨闰荣[2] 隋 雪**[1]
（[1]辽宁师范大学心理学院，大连，116029）（[2]大连大学教育学院，大连，116622）

摘 要 回顾字母文字研究发现，词间空格是有效的词切分线索。而中文没有词间空格，那么是什么线索起着字母文字中空格一样的作用呢？总结以往研究发现，汉字的统计属性、词汇可能性、词素位置、词素熟悉性等因素可能是中文阅读中的词切分线索。字母文字词切分依据物理线索，中文词切分可能依据“心理线索”。未来研究应该解决中文词切分的核心线索以及线索替代的可能性，并构建中文眼动控制模型来解释中文阅读过程及词切分问题。

关键词 词切分 词汇识别 阅读 眼动研究

1 引言

阅读是从书面文字材料中获取相关信息的一种认知活动，是快速而有效获得信息的手段，这个过程从辨识字和词开始，到理解篇章意义为止。其中，词汇识别是关键环节。要进行词汇识别，首先要识别文本中的语言符号，然后把这些符号组合成不同层次的语言学单位，如词、词组、短语和句子等，并确定各单位间的关系，这个过程就是词切分（word segmentation）（Li, Rayner, & Cave, 2009; Liu, Li, Lin, & Li, 2013; Ma, Li, & Rayner, 2014; Rayner, Fischer, & Pollatsek, 1998）。

大多数的字母文字，如英语、西班牙语等，在其书写形式中，词与词之间存在空格，而空格作为一种清晰的物理线索，具有标记词边界的作用，使读者在视觉感知阶段就完成了词切分。词间空格在很大程度上促进了词汇识别，提高了词汇加工速度，删除英文文本中的空格或者用随机字母填充英文文本中的空格会严重干扰阅读理解（Perea & Acha, 2009a; Winskel, Radach, & Luksanneeyanawin, 2009）。

中文是一种没有空格，以字为基本单位的语言，与字母文字不同，中文文本中没有明确的视觉标记，字与字之间间隔大小也相同。每个汉字既可以单独成词，也可以是一个词语或词组的组成部分（Zang et al., 2016），这需要根据语境判断。中文存在很多多义词，一个词的意义也并不唯一。这就使中文词切分的问题变得更加复杂。但是，这些并没有影响正常的汉语阅读，说明汉语阅读中有其特殊的词切分依据，起着与字母文字中的空格一样的作用。在汉语词汇识别过程中，读者很可能会因缺乏词边界信息而难以进行词汇识别，或者因为词切分的方法不同而产生歧义。能否进行准确的词切分直接地影响着汉语阅读。所以，阅读中词切分的研究，有助于揭示阅读的本质，理解阅读的基本现象及规律，也有助于构建汉语阅读眼动控制模型。本文对以往词切分的相关研究进行回顾，并比较字母文字与中文词切分的研究，也对未来词切分的研究进行了展望。

2 词切分的相关研究

字母文字的研究表明，空格是词切分的依据（李玉刚，黄忍，滑慧敏，李兴珊，2017；Bai, Yan, Liversedge, Zang, & Rayner, 2008; Perea & Acha, 2009b），促进词汇识别（Bai et al., 2008）；空格也有助于读者眼跳目标的选择和确定（白学军，李馨，闫国利，2015；Perea et al., 2009b; Winskel et al., 2009）。人为删除词间空格或用字母、数字替代词间空格，均会降低阅读速度，使阅读变得困难，还会导致读者的眼动模式出现实质性的变化，

* 本研究得到国家留学基金委公派留学项目（201608210288）、辽宁省自然科学基金项目（201602460）、辽宁省教育厅一般科研项目（L201683659）和辽宁省教育科学“十三五”规划年度课题（JG16CB341）的资助。

** 通讯作者：隋雪。E-mail:suixue88@163.com

DOI:10.16719/j.cnki.1671-6981.20180511

干扰眼跳在目标词间的空间位置分配（Perea et al., 2009b）。那么中文阅读中，什么线索与空格的作用一致？当去掉此线索时，结果是否与字母文字中去掉空格的效果一致呢？中文中加入空格和字母文字中加入中文的词切分线索的效果又是否相同呢？

2.1 词切分的线索

2.1.1 文本中插入空格

英文的词间空格起到了词切分的作用，那么中文插入空格会不会促进阅读呢？早期的研究并未发现词间空格对中文阅读有促进作用，甚至可能会对正常阅读产生干扰（李馨，白学军，闫国利，臧传丽，梁菲菲，2010），但是后来的研究发现了不一致的结果。Bai 等（2008）研究表明，词间空格对被试的阅读过程没有产生干扰，而非词空格和字间空格延长了被试的阅读时间。沈德立等（2010）的实验结果与 Bai 等（2008）的研究结果一致，空格在中文文本中的作用依赖于空格的位置。空格作用的时间进程研究发现，空格对早期词汇的识别有促进作用（Zang, Liang, Bai, Yan, & Liversedge, 2013）。而 Liu 和 Li（2014）的研究表明，词前空格延长了词加工的时间，但词后插入空格有助于被试的词切分。

从总体上看，汉语中插入字间空格或非词空格没起到有效词切分线索的作用，而词间空格有助于词切分正确地进行。这些研究仅仅回答了空格对汉语阅读的影响，没有回答中文阅读中的词切分依据是什么。当空格作为词切分线索时，被试虽然可以正常阅读，但并没有提高阅读效率。而且空格的存在使文本的排版变得分散，与中文正常的排版区别很大。字母文字中的词切分依赖于空格这样的物理线索，而中文文本却没有明显的物理线索，中文的词切分很可能是依据一种看不见的“心理线索”进行的，不过，这样的推测需要实验进行验证。

2.1.2 文本中插入灰条和改变文字物理属性

Bai 等（2008）认为，使用插入词间空格的方法不能在眼动数据上观察到中文词切分的作用。因为词间空格会导致句子空间分布不同。因此，为了确保句子的空间分布，采用了插入灰条和改变文字物理属性（如亮度、字体粗细、颜色等）的方法来代替空格进行词边界标记（Perea, Tejero, & Winskel, 2015；Perea, Winskel, & Ratitamkul, 2012；Rayner et al., 1998）。当文本中插入灰条时，结果与插入空格条件的结果趋于一致（白学军，郭志英，顾俊娟，曹玉肖，闫国利，2011；Bai et al., 2008）。当改变文字的颜色时，研究结果表明其能帮助读者进行词切分。这从侧面说明空格不是唯一线索。例如三种呈现形式：I am a student from China.（正常文本）、IamastudentfromChina.（去掉词间空格并改变词颜色，例如：“I”是红色字体，“am”是蓝色字体，以此类推）、IamastudentfromChina.（仅去掉词间空格）。此例中，被试仅可以顺利阅读前两个句子。说明空格和颜色这两种物理线索都有助于完成词切分，而被试却难以理解缺乏物理线索引导的第三句话，这说明字母文字的词切分依赖于文本的物理线索。但在中文的阅读中尚不清楚哪些因素起着与字母文字中空格一样的作用，也不清楚哪些线索能够代替它来完成中文文本的词切分。

综上所述，不论对原始材料进行何种改变，人为的词切分都在一定程度上对阅读产生了积极的作用，这与早期研究结果不太一致，原因大致如下：使用了更为精确的仪器；区分了被试的阅读水平，采用不同阅读水平的被试作对比；与以往实验相比，设计更为严格等。

另外，上述都是静态的方法，不能考察词切分的发生位置和阶段，难以确定词切分的数量。张智君等（2012）采用呈现随眼动变化（eye-movement contingent display）技术，通过改变有关词的颜色，提供正确或错误的词边界信息，以此达到协助或干扰词切分的目的，研究汉语阅读过程中词切分发生的位置与数量。结果表明，阅读中文文本时，有两种词切分的方式，一种是加工词 n 导致的，另一种是加工词 n+1 导致的。

2.2 中文正常文本词切分的线索

由于汉语特有的书写方式，中文读者并不依靠空格进行词切分，虽然插入词间空格对中文阅读会有一定的促进作用，但这并不符合中文一贯的书写形式。因此我们需要探讨在正常文本中的词切分依据。

有研究表明，中文汉字的统计属性对中文的词汇识别和眼动控制很重要，读者能够使用统计线索帮助进行词切分和识别（Yen, Radach, Tzeng, & Tsai, 2012）。这说明汉字的统计属性是词切分的有效线索。Zang 等（2016）探讨了词汇可能性在词切分过程中的作用，结果发现，词汇可能性信息对词切分有显著影响，并且影响副中央凹加工的范围。可见，词汇可能性也是词切分的依据。梁菲菲，王永胜和白学军（2016）的研究表明词素位置在某种程度上也是一种词切分线索。梁菲菲等（2016）的另一项

实验发现词素熟悉性可能也是一种词切分的线索。

可见，汉字的统计属性、词汇可能性、词素位置、词素熟悉性等因素可能是中文阅读中的词切分线索。这些线索与空格一样也有着明确的客观标准，但并不是依据肉眼可见的物理线索进行词汇切分的（如空格）。

在阅读空格作为词边界信息的文本时，读者不受年龄、文化程度等因素的影响，由空格引导进行眼跳目标的选择，依据物理线索完成词汇切分（白学军等，2014；李玉刚等，2017；梁菲菲，王永胜，张慢慢，闫国利，白学军，2016；梁菲菲等，2016）。而当汉字的统计属性、词汇可能性、词素位置、词素熟悉性等因素作为词切分线索时，这些线索在文本中很难被去掉，并且与词汇紧密结合（梁菲菲等，2016），由此推测中文词切分的过程可能更加地依赖读者的心理品质，本文用“心理线索”来表示这类线索。除此之外，还有词性、句子成分等也可能是词切分的线索，这需要实验数据的支持。

3 眼动控制模型对词切分的解释

眼动控制模型主要考察的是阅读中的语言因素和眼动生理控制因素如何相互作用，进而产生实际的眼动。近年来，研究者探讨的重要问题是读者在阅读过程中的眼动行为特点，即注视点的位置和注视时间等。

3.1 字母文字阅读的眼动控制模型

有关字母文字的眼动控制模型主要有 E-Z 读者模型（E-Z reader model）和 SWIFT 模型（saccade-generation with inhibition by foveal targets）（陈庆荣，邓铸，2006；Engbert, Longtin, & Kliegl, 2002；Reilly & Radach, 2006）。由于字母文字书写中存在空格，空格标识的边界信息是决定下一次眼跳计划的主要因素（Rayner et al., 1998）。E-Z 读者模型认为，在阅读中驱动眼睛向前移动的主要动力是对词的编码，眼睛接下来注视何处是由文本中低水平的视觉线索决定的，而词汇加工的难易程度决定何时移动眼睛（Perea et al., 2009b）。PVL（preferred viewing location, 即偏好注视位置）通常是由低水平的视觉因素决定，而不是高水平的词汇因素（Winskel et al., 2009）。但是中文文本没有空格，眼睛移动的时间和去向很难用这两个模型准确解释。不排除词汇本身难度对注视时间的影响，但是下一次注视的位置与词切分密不可分，很难用低水平的视觉因素解释。

关于注意的分配方式，E-Z 读者模型主张序列加工，即视野中的文本按照顺序依次进行识别，前一个词没有加工完，后一个词不会进行深度加工。而 SWIFT 模型则主张平行加工，认为注意由词语的激活水平所引导。最新研究表明，在汉语阅读的过程中，注意是基于对象分配的，即注意是一个词接着一个词的序列加工，而不是同时注意多个词的并行方式（Liu & Reichle, 2017）。

3.2 中文阅读的眼动控制模型

与字母文字发展出的多个模型相比，中文的眼动控制模型还在发展阶段。吴俊，莫雷和冷英（2008）通过研究对比，选择 E-Z 读者模型为中文眼动控制模型的基础模型。在此基础上，建构中文词汇切分的策略。细分了英文眼动控制模型里的预加工时间，将中文词切分的影响纳入预加工时间里，使之能够解释中文阅读中的眼动数据。Li 等人（2009）构建了“中文词切分和识别的计算模型”。他们认为，词汇识别与词切分是统一的过程。阅读过程中，视野中字的加工是并行的（Yan, Zhou, Shu, & Kliegl, 2015），而词汇识别是一个串行的过程（Li et al., 2009）。该模型认为词汇识别是由左到右的，视野左侧的词汇有先被加工的优势，下一个词的加工在左侧词汇加工完成之后才会开始，这和 E-Z 读者模型的观点相似。然而，Ma 等（2014）的实验结果却与竞争假说一致，即能够组成的所有词语同时被激活和加工，如果第二个词的词频高于第一个词，则能够改变词的切分形式，不再遵循左侧词汇的加工优势，这说明词切分的结果依赖语境和语义等信息。此模型能够为解释中文词切分的问题和构建眼动控制模型提供一种思路，但是还需要进一步的修改和完善，综合近年来词切分的研究成果，应该着重考虑以下两个方面的问题：字词加工模式和眼跳落点位置。

字母文字的眼动控制模型是建立在“阅读理解是以词为基础”的条件上，中文阅读的研究也证明了词加工的整体性，但这并不意味着字母文字的眼动模型同样适用于中文。根据汉字本身的独特性，汉字成词的多样性和复杂性，在考虑以词语为基本加工单元的同时，也要考虑单个汉字的加工。这就需要探讨汉语阅读过程中的字词加工模式。关于字词加工模式，有以下三种假设：向上反馈假设（feed-forward hypothesis）、整体假设（holistic hypothesis）和交互激活假设（interaction activation

hypothesis）。刘志方，张志君和杨桂芳（2016）的研究结果支持交互激活假设，词语加工影响着单字加工，字词交互激活普遍存在。

关于中文阅读中眼跳落点位置的早期研究发现，首次注视落点的位置平均分布在字的各个位置。而Yan, Kliegl, Shu, Pan, Zhou（2010）的研究认为，在副中央凹存在一个词切分过程，中文阅读者会利用一些从副中央凹获取的信息进行词切分，眼跳落点位置由词切分的结果引导。如果在副中央凹区域能够完成切分，读者的下一次注视落点位置则在词的中央；如果不能完成，则落点位置在词首。张智君等（2012）对词切分数量和位置的研究发现，存在发生于词 n 的加工进程中和发生于词 n+1 的加工进程中的两种词切分现象。这一结果与 Yan 等（2010）的研究结果在是一致的。

也有研究与他们的结果不一致。Li, Liu, Rayner（2011）考察了中文读者的眼跳落点位置，实验结果表明不同词长条件下的 PVL 曲线几乎是相同的。他们认为，中文阅读中的眼动控制不像英文中那样以词的中间位置作为眼跳的目标，词的属性影响对眼跳时间的决定，词切分过程中读者会根据情况选择词的属性或者字的属性进行眼跳目标计划。Li, Liu, Rayner（2015）的研究发现，当眼跳落点临近词尾时，副中央凹能够加工更多的信息。这表示中文阅读者动态的调节眼跳的长度来提高阅读的效率。他们的研究还发现预视的有效性和词频与它的交互作用影响了目标词的眼跳长度，这说明副中央凹的加工对眼跳的落点位置有决定作用。但是，他们也提出眼跳落点和副中央凹的加工并没有关系，因为词的切分和识别是同时发生的。在词没有完成识别的情况下，自然也不会被切分。而 Yan 和 Kliegl（2016）研究了歧义词的词边界对眼跳目标选择的影响，结果显示，在中文阅读中，副中央凹处的词切分影响眼跳的产生，并且验证了高水平的信息参与决定下一步的注视这一假说。

综上所述，对中文阅读的研究发现，中央凹和副中央凹处词的边界信息、词频、语义等信息都影响眼跳的落点，而副中央凹处文本加工的难易程度对眼跳的落点位置起决定作用。

4 未来研究展望

中文词切分过程及规律的研究有助于理解中文阅读过程及中文眼动控制模型的完善。目前的研究取得了很多成果，但也存在一些需要未来关注的问题。

4.1 词切分的线索问题

迄今为止没有文献表明中文词切分存在如空格一样的词切分核心线索。但是中文读者能顺利完成词切分，其词切分的依据是什么？值得未来研究深入探讨。如果存在一个类似“空格”这样的词切分核心线索，能否去掉这个核心线索呢？如果可以，那么去掉之后，是否与字母文字去掉空格的结果一样，导致阅读难以进行呢？除此之外，下划线、字体大小写等这些线索可以代替空格引导词切分，那么中文中的词切分是否也存在替代线索？这些线索能否替代核心线索进行词切分？前文表明，汉字的统计属性、词汇可能性、词素位置、词素熟悉性等因素可能是中文阅读中的词切分线索，那么哪一个可能是核心线索？这些问题需要未来的实验研究进行证明。

4.2 词切分的内在机制问题

鲜有文献提及，在字母文字的正常阅读中，负责词切分的脑区以及词切分过程中的脑部活动。阅读过程中，不能正确进行词切分，导致阅读效率低，甚至理解困难，其对应的脑机制是什么？正确和错误的词切分在哪些脑区会出现不同反应？时间进程又有什么差异？中文正常文本阅读中，词切分的依据明显不同于有空格的字母文字，那么，正确和错误词切分相关的脑区及活动与加工字母文字时是否一样？希望未来研究能解决这些问题。

另外，当删除空格时，字母文字的词切分变得困难，这种词切分状态，脑部活动发生了怎样的改变？如果中文词切分的线索能够进行控制，那么，词切分线索改变导致加工困难时，脑部活动又将会发生什么改变？

让读者加工两种词汇，一种是只有一种切分可能的词汇（如：老一辈，属于 A+B+C 类型），另外一种是有两种切分可能的词汇（如：资本家，属于 AB+C 类型；总司令，属于 A+BC 类型），记录词汇加工的脑部活动情况，也许会发现不同的脑机制。

另外，词切分是否是副中央凹预视的结果，也有待研究。当采用边界范式控制有效预视时，词切分受到的影响及词切分结果的变化值得思考。苏衡，刘志方和曹立人（2016）的研究表明中文阅读者对词 n+1 的切分是通过词汇预测完成的。而预测存在失误的可能，如何发现并调整失误也值得探究。

参考文献

白学军，郭志英，顾俊娟，曹玉肖，闫国利 . (2011). 词切分对日－汉双语者汉语阅读影响的眼动研究 . *心理学报，43(11),* 1273–1282.

白学军，李馨，闫国利 . (2015). 汉语阅读眼动控制：20 年研究的总结 . *心理发展与教育，31(1),* 85–91.

白学军，张慢慢，臧传丽，李馨，陈璐，闫国利 . (2014). 词边界信息在中文词汇学习与识别中的作用：眼动研究的证据 . *心理科学进展，22(1), 1–8.*

陈庆荣，邓铸 . (2006). 阅读中的眼动控制理论与 SWIFT 模型 . *心理科学进展，14(5),* 675–681.

梁菲菲，王永胜，白学军 . (2016). 词间空格促进汉语阅读的新词学习：词素位置的调节作用 . *心理学探新，36(5),* 403–408.

梁菲菲，王永胜，张慢慢，闫国利，白学军 . (2016). 新词词素熟悉性影响视觉词切分线索在新词学习中的促进作用 . *心理科学，39(2),* 258–264.

李馨，白学军，闫国利，臧传丽，梁菲菲 . (2010). 空格在文本阅读中的作用 . *心理科学进展，18(9),* 1377–1385.

李玉刚，黄忍，滑慧敏，李兴珊 . (2017). 阅读中的眼跳目标选择问题 . *心理科学进展，25(3),* 404–412.

刘志方，张智君，杨桂芳 . (2016). 中文阅读中的字词激活模式：来自提示词边界延时效应的证据 . *心理学报，48(9),* 1082–1092.

沈德立，白学军，臧传丽，闫国利，冯本才，范晓红 . (2010). 词切分对初学者句子阅读影响的眼动研究 . *心理学报，42(2),* 159–172.

苏衡，刘志方，曹立人 . (2016). 中文阅读预视加工中的词频和预测性效应及其对词切分的启示：基于眼动的证据 . *心理学报，48(6),* 625–636.

吴俊，莫雷，冷英 . (2008). 中文词汇切分对眼动数据的影响 . *心理学探析，28(2),* 41–48.

张智君，刘志方，赵亚军，季靖 . (2012). 汉语阅读过程中词切分的位置：一项基于眼动随动显示技术的研究 . *心理学报，44(1),* 51–62.

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1277–1287.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42(5),* 621–636.

Liu, P., & Li, X. (2014). Inserting spaces before and after words affects word processing differently in Chinese: Evidence from eye movements. *British Journal of Psychology, 105(1),* 57–68.

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

Li, X., Liu, P., & Rayner, K. (2015). Saccade target selection in Chinese reading. *Psychonomic Bulletin and Review, 22(2),* 524–530.

Liu, P. P., Li, W. J., Lin, N., & Li, X. S. (2013). Do Chinese readers follow the national standard rules for word segmentation during reading? *PLoS ONE, 8(2),* e55440.

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

Liu, Y., & Reichle, E. D. (2017). Eye-movement evidence for object-based attention in Chinese reading. *Psychological Science,* 956797617734827.

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. *Journal of Experimental Psychology Human Perception and Performance, 40(3),* 1046–1059.

Perea, M., & Acha, J. (2009a). Space information is important for reading. *Vision Research, 49,* 1994–2000.

Perea, M., & Acha, J. (2009b). Does letter position coding depend on consonant/vowel status? Evidence with the masked priming technique. *Acta Psychologica, 130(2),* 127–137.

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research, 38,* 1129–1144.

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Minds Eye, 7(1),* 34–55.

Perea, M., Tejero, P., & Winskel, H. (2015). Can colours be used to segment words when reading? *Acta Psychologica, 159,* 8–13.

Perea, M., Winskel, H., & Ratitamkul, T. (2012). On the flexibility of letter position coding during lexical processing: The case of Thai. *Experimental Psychology, 59(2),* 68–73.

Winskel, H., Radach, R., & Luksanneeyanawin, S. (2009). Eye movements when reading spaced and unspaced Thai and English: A comparison of Thai-English bilinguals and English monolinguals. *Journal of Memory and Language, 61,* 339–351.

Yan, M., & Kliegl, R. (2016). Carprice versus carprice: Word boundary ambiguity influences saccade target selection during the reading of Chinese sentences. *Journal of Experimental Psychology: Learning Memory and Cognition, 42(11),* 1832–1838.

Yan, M., Kliegl, R., Shu, H., Pan, J., & Zhou, X. (2010). Parafoveal load of word n+1 modulates preprocessing effectiveness of word n+2 in chinese reading. *Journal of Experimental Psychology Human Perception Perform, 36(6),* 1669–1676.

Yen, M. H., Radach, R., Tzeng, O. J. L., & Tsai, J. L. (2012). Usage of statistical cues for word boundary in reading Chinese sentences. *Reading and Writing, 25,* 1007–1029.

Yan, M., Zhou, W., Shu, H., & Kliegl, R. (2015). Perceptual span depends on font size during the reading of Chinese sentences. *Journal of Experimental Psychology: Learning Memory and Cognition, 41(1),* 209–219.

Zang, C., Liang, F., Bai, X., Yan, G., & Liversedge, S. P. (2013). Interword spacing and landing position effects during Chinese reading in children and adults. *Journal of Experimental Psychology: Human Perception and Performance, 39(3),* 720–734.

Zang, C., Wang, Y., Bai, X., Yan, G., Drieghe, D., & Liversedge, S. P. (2016). The use of probabilistic lexicality cues for word segmentation in Chinese reading. *The Quarterly Journal of Experimental Psychology, 69(3),* 548–561.

The Basis of Chinese Word Segmentation and Explanation of Eye Movement Control Model

Xiang Huiwen[1], *Hu Xinzhuo*[1], *Yang Runrong*[2], *Sui Xue*[1]
([1]Department of Psychology, Liaoning Normal University, Dalian, 116029)
([2]Department of Education, Dalian University, Dalian, 116622)

Abstract In the process of reading, word recognition is the key for readers to acquire effective information. In the process of word recognition, we must first identify the language symbols in the text, and then combine these symbols into different levels of linguistic units, such as words, phrases and sentences, and to determine the relationship between the various units. This process is called word segmentation. In the writing system of most alphabetic languages, such as English, German and Spanish, there are spaces between words, and spaces act as a clear physical clue to mark the word boundary. Moreover, spaces between words provide readers with clear visual information so that the readers complete the word segmentation in the visual perception stage, and the probability of word segmentation errors is very low. Reviewing the alphabetic languages studies, we found that the space between words is valid for word segmentation. Space can help readers to determine the boundaries between words, and it can also promote word recognition and improve word processing speed. Nevertheless, Chinese is a kind of language based on characters. Different from any alphabetic language, there is no clear visual mark between words in Chinese text, and the spacing between words is the same. Each Chinese character can be either a single word or as part of a word or phrase. Depending on the context of the sentence, a Chinese character can be either an independent word or a part of a word. There are many ambiguous words in Chinese, which depends on the context to determine its meaning, and then determine whether it is an independent word or a part of the vocabulary, which makes the Chinese word segmentation more complicated. However, these do not affect the normal Chinese reading, indicating that Chinese reading must have its own special word segmentation basis, which plays the same role as the spaces in alphabetical languages. Because there is no space between the Chinese characters, it should be discussed what can be the same as the space in English to help Chinese readers to segment words. In the previous studies, it has been found that the statistical attributes of Chinese characters, vocabulary possibilities, morpheme position, morpheme familiarity and other factors may be the mark of word segmentation in Chinese reading. A latest research suggests that in the reading process, the text is read in a strictly continuous (perhaps object-based) way. As we know, the word segmentation of alphabetic languages is according to physical clues, such as the space between words, but the word segmentation in Chinese may be based on psychological clues. The process of word segmentation is also related to other factors, such as individual differences and preview effect. This article holds that the future research should resolve the core clues of Chinese word segmentation and the possibility of clue substitution, and construct Chinese eye movement control model to explain the Chinese reading process and word segmentation problem.

Key words word segmentation, word recognition, reading, eye movement research