

# 儿童外语学习认知数据收集的在线游戏框架\*

马为之, 张敏, 张琛昱, 陈忆馨, 谢倩, 孙炜岳, 刘奕群, 马少平  
(智能技术与系统国家重点实验室, 清华信息科学与技术国家实验室(筹), 清华大学计算机系,  
北京, 100084)

z-m@tsinghua.edu.cn

**摘要:** 近年来, 人工智能技术飞速发展, 不少工作试图从人类的认知发展过程中探索前进方向, 语言学习认知的过程成为了重点关注的研究领域。已有的语言认知研究工作主要集中在学龄前儿童母语的词汇学习认知方面, 依赖于 WordBank<sup>1</sup>等大规模语料库。然而就我们所知, 目前在第二语言学习方面研究不多, 尚未有大规模的第二语言词汇学习数据, 且传统的数据收集方法难以收集到大规模数据, 这也一定程度上限制了对于第二语言学习的研究工作及母语与第二语言学习的比较。针对这一问题, 本文针对学龄前儿童群体设计了基于游戏性原则的数据收集方法和研究框架, 用于收集第二语言的语言学习情况和用户数据, 以支撑相应研究工作的开展。目前, 我们已经实现了针对学龄前儿童的第二语言为英语的词汇认知数据收集系统, 正在进行在线的数据收集。

**关键词:** 语言认知; 儿童; 第二语言; 游戏性数据收集。

中图分类号: TP391

文献标识码: A

## An Online Data Collecting Framework by Game with a Purpose for Children Second Language Development Research

Weizhi Ma, Min Zhang, Chenyu Zhang, Qian Xie,  
Weiyue Sun, Yiqun Liu, Shaoping Ma

(State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

**Abstract:** In recent years, the technology of artificial intelligence has developed rapidly, but it is still far from human intelligence. Many studies try to explore further improving method from the research of human cognition. In conducted work, the language cognition research is often based on the dataset of children's first language vocabulary development, such as: WordBank and other large-scale corpora. However, there is no large-scale second language vocabulary development dataset, and it is very difficult to collect a big dataset with traditional data collecting method. This limits the study of second language learning and the comparison of first language and second language learning. In this paper, we designed a data collecting framework for children based on the idea of games with a purpose, to collect children's vocabulary development status and his/her attributes. We have implemented the second language vocabulary development collecting system for children English learning so far, and the system is conducting online data collection now.

**Key words:** Language Cognition; Children; Second Language; Games with A Purpose.

## 1 引言

随着计算机计算能力和互联网产生的数据规模的提升, 以深度学习为代表的人工智能算法在近年取得了越来越好的表现, 特别是在文本处理、图像识别等方面都有了很大的提升。

---

\* **基金项目:** 本文工作得到国家自然科学基金(61532011, 61672311)资助

**作者简介:** 马为之(1992-), 男, 博士生, 主要研究用户画像、个性化推荐; 张敏(1977-), 通讯作者, 女, 副教授, 主要研究为信息检索与挖掘、个性化推荐与用户建模。

<sup>1</sup> <http://wordbank.stanford.edu/analyses>

然而，目前人工智能的发展水平与人类智能仍有很大差距。已有的研究工作对于人类的知识学习和认知等方面的研究和探索还不够深入，我们对于人类的认知过程了解还很初步，近年来也有更多的工作致力于人类的认知学研究，希望能够从中找到前进的方向。

语言学习认知是人类学习认知的重要部分。虽然这一过程受个体特征、个体的成长环境等社会学因素和个体教育情况的影响，但是语言知识本身具有一定的架构性和规则性，例如：存在语法规则和基本的组成单位；以字/字母为最小粒度的符号；以拼音/音标为最小粒度的发音等。人类在语言学习过程中受到这些规则的限制，且学习同一语言的不同个体学习目标相同。因此，研究语言学习认知的过程也成为了研究人类认知的重要手段。实际的研究过程中，往往可以通过控制变量和大规模调研，在不同个体之间进行语言学习认知结果的比较，更可以从语言学习的过程中，总结人类的认知规律，帮助识别人类的学习特性。可以说，从语言学习认知的角度去尝试了解人类的学习认知过程和知识构建过程是较好的研究切入点。

然而，人类的各个年龄阶段都可能存在语言学习认知的过程，从婴幼儿时期的母语学习模仿，到成年后的外语学习，单纯从语言学习认知过程进行认知研究将面临变量太多、因素不可控等问题；因此不少工作将研究重点放在人类早期，特别是学龄前儿童时期的语言学习认知过程[1-3]。已有工作中，语言词汇学习情况是研究语言学习与认知的重要根据，如：1996年由 Fenson 等人提出的针对英文的 CDI 词汇量表，被广泛应用于英语学习认知过程的研究[4]。大部分已有工作以母语为语言学习的研究目标[1,2,3,5]，然而，本文更倾向于将第二语言作为语言学习认知研究的载体：一方面，我们认为母语的学习研究将更可能受到复杂环境的影响，特别是对于婴幼儿的语言学习过程，虽然他们开始模仿语言并开始有一定的理解能力，但是限于年龄发展水平，我们很难获得他们的直接反馈；另一方面，我们也希望能够进行第二语言的学习认知研究，可以与母语的一些研究结果进行对比，从而更深层次地比较语言学习发展的过程。

由于语言学习认知这一研究问题具有开放性的特点，研究过程中需要大量的数据统计结果作为支撑，而目前缺乏已有的大规模数据用于第二语言学习规律的发现与建模，更难以进行母语学习和第二语言的学习过程比较。本文致力于首先解决这一基础性问题，为进一步的深入研究提供数据支撑，本文的主要贡献如下：

1. 据我们所知，这是首个针对中国学龄前儿童的大规模第二语言外语学习认知过程的研究工作，我们明确了研究方向和需求，确立了在线收集词汇认知数据的思路，为第二语言的学习研究与与母语学习的认知过程比较打下了基础；
2. 结合语言学和心理学等学科知识，我们设计了游戏性数据收集流程，能够用于收集大规模的语言词汇认知数据及被试基本情况，并且实现了以英语为第二语言的数据收集系统；
3. 整个框架设计不局限于某种语言的应用，可以广泛应用于设计不同语言的学习认知发展的数据收集系统。

## 2 相关工作

### 2.1 针对语言学习词汇认知的相关研究

在语言学习认知的研究中，词汇认知水平是评价语言学习情况的重要标准，它能够被明确的定义和比较。其中，由 Fenson 等人在 1996 年提出的 CDI(The MacAuthor Communicative Development Inventories)词汇量表，对于近些年语言学习认知的研究起到了重要的作用[4]。

CDI 是一套用来测量婴幼儿和早期学龄前儿童语言发展水平的测试量表，具有高效、经济的特点。同时，克服了其他测量手段的局限性——不需要经过专业的训练便可以使用、不需要严格的条件限制等。该量表自提出后经过了长时间的数据统计与检验，被证明可以有效

地体现学龄前儿童的真实语言水平。由斯坦福大学发起的 WordBank 是利用该量表的数据建立起的世界儿童语料数据库，目前已经各国语言的母语 CDI 量表结果共计数十万条记录在库，为心理学、语言学、生物学等学科的发展提供了宝贵的研究资料[1-3]。

在词汇认知规律研究的基础上，有部分工作尝试了对人类的语言学习过程和词汇网络模型进行刻画，Nicole 与 Eliana 整理了已有的针对人类大脑中词汇网络构建和认知模型的研究工作[6]；Nelson 等人根据针对数千人的面对面调查，提出的语义网络的构建方式[7]；Hills 等人根据 CDI 的词汇研究结果，总结了人类学习新词语顺序的模型[5]。可以说，大规模数据是深入语言认知研究的重要助力。

## 2.2 针对第二语言学习的相关研究

WordBank 中已有的 CDI 量表数据是通过统计学龄前儿童日常使用母语的情况得到的，目前并没有针对以英语不为官方语言地区的英语作为第二语言学习的相关数据，这使得已有的大量研究工作是基于母语的认知学习研究，或者是不同地区的母语语言认知学习过程研究。对于第二语言学习的相关研究工作较少，但我们认为研究第二语言学习过程有着重要意义。已有的不少针对第二语言的学习认知过程以推理和假说为主，试图解释第二语言学习与第一语言学习的异同，但是不少缺乏相应的大规模数据用于支撑其研究结果[8]，这其中也包括不少研究中国学龄前儿童学习英语过程的研究[9,10]。

同时值得注意的是，大部分语言认知的研究会以学龄前儿童为目标研究群体，这是因为人类很早就开始接触语言，而儿童时期的语言学习认知研究变量更可控，且这段时间人类的接受能力和学习能力很强。但如前所述，目前儿童的第二语言词汇学习认知情况的大规模数据稀缺，因此，设计和创建儿童第二语言的学习数据库对于相关研究有着非常重要的意义。

## 2.3 游戏性数据收集工作

大规模的数据收集，特别是需要人工标注确认的数据，如果采用雇佣标注的方式进行，往往需要耗费大量的人力和物力。因此，近年来涌现了一批利用网络用户实现数据高质量标注的游戏性数据收集工作，这类工作通过良好的流程和游戏性设计，使得用户参与到标注过程中，同时也得到放松，实现了数据收集者和标注者的双赢。代表性工作有：Luis 提出的 reCAPTCHA 字符识别系统，将难以识别的字符作为验证码让用户识别输入，每年可以识别上十亿条内容且正确率高达 99.1%[11]；此外，他与 Shiry 等人还共同设计了 ESP game[12]、Phetch[13]和 Peekaboom[14]三个游戏性数据收集系统，以 ESP game 为例，它通过让两个不同的用户对于同一幅图片找出他们观察到的物体，完成对于图片物体的标注，而用户在参与过程中则享受到了游戏的快乐，因而有动力主动完成更多的标注任务。

受到这些基于游戏性设计数据系统的启发，我们考虑将学龄前儿童的语言学习过程中的词汇认知水平也设计成类似的系统，能够让参与者从中获得激励和快乐，以实现用户主动参与数据收集过程，从而不需要耗费大量的人力物力进行数据的收集工作。

# 3 研究目标分析

本文致力于设计并实现切实可靠的大规模数据收集系统，着重考虑研究人类早期第二语言学习认知发展过程的词汇学习和网络构建过程。目标研究群体为 3-6 岁的儿童，重点研究他们的第二语言的词汇学习过程。结合研究目标和已有相关工作，本文首先进行了如下分析：

## 3.1 设计原则

在进行数据收集系统的设计过程中，我们将遵循以下几条设计原则：

- **无害性原则**：面向学龄前儿童的数据收集过程，不应当影响和危害儿童的身体和心理健康，也应当避免过程中对儿童产生负面反馈，避免影响其健康成长；
  - **有效性原则**：要保证数据收集过程和结果的有效性，特别是针对儿童的词汇认知情况收集过程，需要保证与儿童有充分良好的沟通，最好能够让家长参与到数据收集的过程中；
  - **科学性原则**：由于语言学习和认知涉及了多门学科，有相对复杂的变量，因此，变量的选定和流程的设计均需要保证科学合理，且应尽量使得数据完整丰富；
  - **匿名性原则**：数据收集过程应当充分尊重儿童的隐私，注意避免收集敏感数据。
- 所有的框架设计和实现工作都将基于这四条原则的基础上开展。

### 3.2 数据收集方式

我们首先面临的问题就是如何设计数据的收集方式，以收集到大规模具体到单词级别的学龄前儿童词汇认知情况用于进一步的研究。而针对语言学习认知的已有研究工作中，数据收集主要有两种方式：

1. 对于学龄前儿童学习过程进行音频、视频记录，然后对内容进行人工分析，了解儿童的学习认知情况；
2. 由数据采集人员与学龄前儿童或者与儿童的家长进行当面交流，通过答题的方式了解相关信息。

然而，这两类采集的方法都有着明显的弊端，这也是数据收集过程中将面临的主要难点：

1. 无论是通过收集语音还是现场交流，都需要耗费大量的人力成本，传统的收集方法收到此因素的限制，很难获得大规模数据；
2. 对学龄前儿童学习的音频、视频全程记录有可能会侵犯到隐私；
3. 学龄前儿童的认知交流能力有限，数据收集人员在沟通方面很可能遇到各种问题，引入家长辅助则有可能影响数据的收集结果。

在这样的情况下，受启发于已有的游戏性数据标注工作，本文设计了基于游戏性的数据收集方式，用于了解学龄前儿童在第二语言学习过程中的语言认知，特别是词汇认知的水平。同时，不同于已有的游戏性数据标注工作，本文设计的收集实验如果仅仅学龄前儿童参与，可能很难确保数据收集结果的有效性，因此，设计的环节将要求家长参与以保证数据质量。

在综合考虑以上困难和我们设计原则的基础上，本文设计的游戏性数据收集方法主体思路如下：以了解儿童的第二语言学习水平和状况为测试的游戏性吸引点，通过移动互联网的界面实现交互，由家长发起实验，对孩子进行词汇认知的问答操作，最后反馈该儿童的第二语言学习水平。一方面，反馈给用户的儿童语言学习水平是收集过程游戏性的产出，也是家长和儿童参与游戏的激励。另一方面，游戏过程中儿童的词汇认知测试结果，则是作为开展相关研究的数据来源。

这样的实验设计能够同时满足数据收集方与参与方的需求：对于数据收集方而言，在实验过程中可以方便地得到详细的儿童对第二语言词汇的认知情况，同时家长的参与令测量结果更有可信度。线上收集的方式也令大规模的数据收集变得不再困难，能够节约大量人力成本；对于参与方而言，家长对于孩子的各方面能力测试会有积极了解的心态，这样的词汇量测试的方式，加上最终能够直接反馈的语言学习水平结果，能够加深他们对于儿童的了解，也为游戏的大规模传播提供了可能性。

### 3.3 问答词汇选取

值得注意的是，即便是采用了游戏性的数据收集方式，在交互过程中，受限于参与者的精力、耐心情况，实际的系统设计不能给予被测试的学龄前儿童及其家长太大规模的词语列

表并要求他们作答。因此，在设计的过程中，需要注意使用尽量少的单词来实现对儿童词汇认知情况的刻画，同时也要保证收集到数据的覆盖面情况。

在相关工作中，基于英文的 CDI 词汇量表是一套非常成熟的词汇量表，WordBank 词库对于学龄前儿童的英语学习词汇发展过程积累了大量的数据资源，有数万儿童的 CDI 词汇量表测试结果。因此，我们希望借助词表中的词语对儿童的词汇认知情况进行相应的预测。以英文的词汇认知研究为例，WordBank 数据集中针对英语母语的幼儿词汇认知情况，统计了 558 个最常用的 CDI 量表中的词语，可以用于母语为英语的词汇认知研究工作。

然而，即便是在第二语言学习的研究过程中仅使用 WordBank 中的 558 个词语，考虑到家长和学龄前儿童的交互耗时，如果一次测验要求检测这些词汇的认知情况时间将会花费被试大量的时间和经历。因此，在设计过程中可以考虑引入难度分级制度，同时将待辨认的词汇分为固定词语（即一定会被要求回答的）和随机词语两种，这样可以保证收集来的数据一定程度上集中，可以用于规律性的研究和发现，同时也可以做个体的学习情况研究。

此外，在选词过程中，为了能够进一步探究学龄前儿童的理解认知能力，我们在根据 CDI 选词的过程中保留母语儿童的词语认知难度，用于比较母语与第二语言学习的认知过程差异；另一方面，在词语的选择方面，也应考虑其覆盖性和种类，如：具体词和抽象词对比等。

### 3.4 影响语言学习认知的可能变量

确立了数据收集方式后，还需要明确的是除词汇认知情况外其他研究过程中较为关注的变量，这些信息应该同时在词汇认知情况收集过程中得到，本文将已知的有可能影响学龄前儿童进行第二语言学习的主要变量归纳为了如下几类：

- 人口社会学变量：学龄前儿童的社会学属性及成长环境情况，包括性别、年龄、受教育情况、家庭经济状况、家庭中交流语言等；
- 语言学习能力：学龄前儿童在语言学习方面的学习能力，已有的测量方式包括短期记忆能力、听力和听觉水平等；
- 非语言智力水平：学龄前儿童对于事物和规律的辨析和发现的能力，以智力测验为代表，被广泛应用的有瑞文推理[15]等度量方式。

这些已知变量和尚未发现的其他变量的存在使得个体间的语言学习认知结果产生了差异，因此，对于这些变量的分析将能够有助于我们了解和归纳不同变量对于语言学习认知过程中造成的正面和负面影响。在实验的框架中，可以考虑收集儿童个体在这几方面的表现情况，用于未来的深度比较分析。值得注意的是，这些变量和认知能力的情况是对于儿童第二语言词汇认知情况的额外信息，考虑到儿童实际参与的时长和兴趣程度，除了基本的信息，如：儿童的性别、年龄外，其他的信息收集均可考虑作为问题体现在收集过程中。

综合这几方面的分析，本文设计了基于游戏性设定的数据收集框架，将在第四章中进行详细介绍，同时也将介绍儿童第二语言词汇认知研究模型的具体实现方法。

## 4 游戏性数据框架及实现

### 4.1 模型框架设计

图 1 展示了基于游戏性设定的词汇测试数据收集流程框架。除了在第三章进行的各方面分析，考虑到数据收集的过程中由于缺乏监督和强制手段，用户随时都有可能中途退出，如果在进行词汇能力测试之前有太多的信息收集环节，将很可能影响用户参与的积极性，我们将框架分为了必选测试内容和可选测试内容两个模块：

必选测试内容在家长与孩子选择参与后，首先收集最基本的信息，如：孩子的年龄、性别及接触第二语言的渠道。随后将开始进行词汇水平测试，在测试过程中将设计词语难度分层和选词策略，用于尽可能多且有效地收集到相关数据。在回答了一定数量（目前为 20 道）的词汇测试题后，系统将会自动生成反馈报告，并引导用户自主参与可选测试内容。必选测试是该数据收集系统的核心部分。

可选测试内容包含三部分，图中的虚线框代表了这一环节是非必须参与的，这三部分对应了在 3.3 章节中提到的三类可能影响语言学习的主要变量，每个模块后用户也会得到相应的结果反馈。这三部分分别是：更详细的用户信息调查问卷（如第二语言的使用频率、孩子接受第二语言教育的方式和频率、家庭基本状况等），用于进一步补充基本信息情况；以瑞文推理测验为代表的非语言智力测验和以短期记忆测验为代表的语言学习能力测验。可选测试的目的是进一步收集测试者个体的各方面情况，用于分析各类变量对于第二语言学习认知的影响，为之后可能的建模和进一步探索提供更加详细的数据支撑。

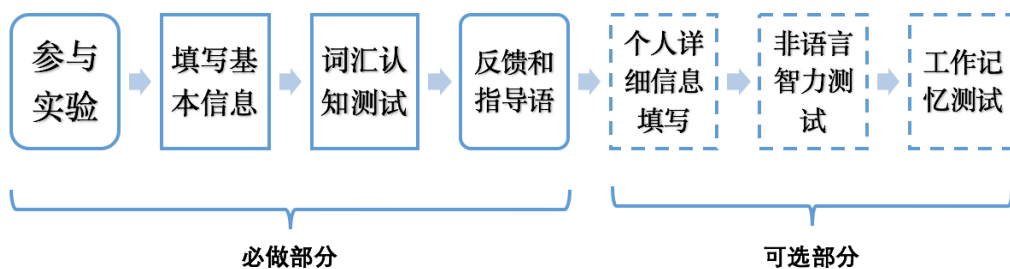


图 1 数据收集框架图

## 4.2 基于中国学龄前儿童英语语言学习认知研究的数据收集框架实现

在 4.1 节框架图的基础上，我们针对母语为中文的中国学龄前儿童的英语学习评价，具体实现该流程。下面将各模块的具体设计介绍如下：

### 4.2.1 词语难度分级及选词策略

如 3.2 节中所述，词语的难度分级对于产生待回答的词语和给予用户评价反馈有重要意义，且由于针对的是英语学习的场景，因此我们沿用了 WordBank 项目中用于儿童英语词汇测验的 558 个 CDI 测试词语。通过比照以英语为母语的 10,000 余名年龄在 16 个月到 30 个月的学龄前儿童对于这五百余词语的掌握情况，我们将这些词语划分到了  $L_1 - L_5$  5 个难度级别。对于每个词语  $w$ ， $f_w(x)$  代表年龄为  $x$  的儿童认识词语  $w$  的比例，找到  $x^*$  满足：

$$x^* = \operatorname{argmin}_x (f_w(x) \geq 0.5)$$

在这个词的难度级别  $w$  为

$$w \in \begin{cases} L_1 & 16 < x^* \leq 18 \\ L_2 & 18 < x^* \leq 21 \\ L_3 & 21 < x^* \leq 24 \\ L_4 & 24 < x^* \leq 27 \\ L_5 & 27 < x^* \leq 30 \end{cases}$$

我们将英文母语的 CDI 测试结果作为初步的难度划分依据，使得我们在收集到实验结果后可以开展母语学习和第二语言学习认知过程的比较工作。最终，得到的  $L_1 - L_5$  这五个难度级别的词语数量分别为 41, 51, 194, 164 和 108。

在每个难度级别  $L_i$  中，我们从中人工选取了 4 个词语作为该等级的固定词。这四个词语将兼顾具象词与抽象词，以第四等级为例，我们选择了 ride、broken、hamburger 和 cloud 四个词语作为固定词语。其他的词语以随机选取的方式出现在用户的评测过程中。因此，难度级别为  $i$  的所有词语将被划分为由 4 个固定词构成的集合  $C_i$  和剩余词集合  $R_i$ 。

此外,为了能够在收集过程中动态调整学龄前儿童会面临的词语难度,主要考虑参与者在回答过程中收益 Gain 的变化情况,我们设计了如算法 1 的难度调整策略:

```
IF  $C_x \neq \emptyset$ :
    # 首先从固定词集中选词,选取后将该词移除集合
     $w = \text{Random}(C_x)$ ,  $C_x = C_x \setminus \{w\}$ 
     $\text{result} = \text{Question}(w)$ 
    # 根据答题正确与否得到 Gain,有可能是正或者负
     $\text{Gain} += (\text{result} == \text{TRUE})? A_w : -A_w$ 
    # 如果固定词集回答完毕,进行评价
    IF  $C_x == \emptyset$ :
        IF  $\text{Gain} > \alpha$ :
             $x += 1$ 
        ELSE IF  $\text{Gain} < \alpha$ :
             $x -= 1$ 
        # 无论是否难度变化,清零 Gain 和 K
         $\text{Gain} = 0, K = 0$ 
    ELSE:
         $w = \text{Random}(R_x)$ ,  $R_x = R_x \setminus \{w\}$ 
         $\text{result} = \text{Question}(w)$ ,  $K += 1$ 
         $\text{Gain} += (\text{result} == \text{TRUE})? A_w : -A_w$ 
        # 回答两道题之后,考虑是否需要跳转
        IF  $K \geq 2$ :
            IF  $\text{Gain} > \beta$ :
                 $x += 1, \text{Gain} = 0, K = 0$ 
            ELSE IF  $\text{Gain} < \beta$ :
                 $x -= 1, \text{Gain} = 0, K = 0$ 
            # 如未满足跳转条件,只保留当前步获得的 Gain 值
             $\text{Gain} = (\text{result} == \text{TRUE})? A_w : -A_w$ 
```

算法 1 选词难度的调整策略

每次要选词时算法 1 都会被运行一次:算法会先从固定词集产生词语,回答所有固定词后将会进行回答初步评分,如果评估结果不足以进行难度调整,则会从剩余词语中随机生成词语继续评价。 $\text{Question}(w)$ 函数表示向用户询问词语  $w$ 。新用户进入系统时候初始的难度等级  $x=1$ ,  $A_w$ 代表词语  $w$  的评分。实际系统中, Gain 和 K 的初始值均为 0,且每个词语的评分都设定为 1,  $\alpha$  和  $\beta$  是两个参数在实际系统中我们将其分别设为了 2 和 1。

在每个单词的测验中,需要家长在手机上播放单词发音<sup>2</sup>,学龄前儿童聆听后将需要从页面中的四幅图片中选择与之最相关的一幅后完成提交。其他三幅图将通过从收集的图片库中采用随机的方法产生,所有的题目都已经经过三轮的人工检验,保证每道题目的正确答案唯一。在收集系统中,我们将每个用户需要回答的词汇题数量定为 25 道。

#### 4.2.2 非语言智力测验及短期记忆测验

在系统实现过程中,我们选取了瑞文标准推理测验[15]和工作记忆测验[16]作为儿童的非语言智力测验及短期记忆测验的具体方式。瑞文标准推理测验是由英国心理学家瑞文于

<sup>2</sup> 单词的音频文件收集自 <http://fanyi.baidu.com>, 美式发音方法。

1938 年创制，在各国沿用至今，它用于测验一个人的观察能力及思维能力，是纯粹的非文字智力测验，所以广泛应用于无国界的智力/推理能力测试。我们从标准的瑞文推理测试中选择了适合儿童的 8 道代表性的题目作为测验内容[17]。

工作记忆测验是认知心理学提出的有关人脑中存储信息的活动方式，工作记忆是指短期的记忆场景下的记忆结果，已有工作的研究结果发现较强的工作记忆能力能够帮助用户的语言学习[18]。通常的工作记忆是通过短时间内记忆正确的最长随机数字的长度来进行测量，成年人平均的记忆长度为 7 左右。本系统中工作记忆测验在韦氏智力测验中的数字广度测验的基础上略作调整，使得测验时间更短，并采用了如下的策略动态调整测验数据的长度。初始的数字长度为  $L = 3$ ：

```
IF 能够在听音后完全复述正确：
    L += 1
ELSE IF 连续复述错两次：
    L -= 1
```

算法 2 工作记忆测验的数字长度调整策略

### 4.2.3 系统反馈内容的设计

由于框架采用了游戏性设计的思路，特别是在反馈过程中需要遵循我们的无害性设计原则，因此，除了考虑数据收集过程中的变量设计和科学性设计外，还需要关注用户反馈内容的设计，以满足测验者的兴趣和使用需求，确保不会对儿童的成长造成负面的影响和打击。这对于数据收集系统的推广有着关键作用。

值得注意的是，在测验过程中我们并不会显示的对于单道题目给予正确或错误的反馈，避免打击参与者的积极性。而在词汇学习评估、瑞文推理测验结果和工作测验结果的反馈内容中，分别采用了以下三种反馈的方法，用于激发和引起参与者的兴趣：

- 词汇学习评估结果的反馈：根据回答者答对的最难单词对应的级别，反馈告知用户内容如下：您家宝贝已经开始掌握 X 个月大的英语母语孩子开始掌握的单词啦！如果出现极端情况，全部答错，则反馈“看来宝贝刚开始接触英语，不妨从简单的日常用语开始学习吧~7 岁以前是儿童学习语言的黄金时期，加油呀！”
- 瑞文推理测验结果：给出孩子答题正确率及完整瑞文推理量表中不同周岁孩子正确率的中位数。
- 工作记忆测验结果：反馈孩子在数字记忆中记住的最长数字长度，并告知普通成年人的平均记忆长度为 7。

系统反馈内容的设计是数据收集系统游戏性的重要体现，也是参与者帮助进行推广的重要驱动力，因此，对于每条反馈都需要进行仔细的斟酌，给参与者以足够的激励。

### 4.3 系统展示

我们根据 4.1 和 4.2 的设计内容，完成了数据收集系统的搭建，开发了适配手机页面的游戏性数据收集接口，图 2-图 5 是目前完成的系统设计的截图，展示了首页、词汇评测页面、瑞文推理测验页面和工作记忆测验页面。

整个系统已经开放用于数据收集，可通过 <http://kidswords.thunlp.org:15010> 访问。我们会将收集到的数据集进行脱敏处理后进行发布，用于支持相关的语言认知发展研究工作。

## 5 结论与未来工作



本文针对语言认知学习过程的研究工作,提出了一个基于游戏性设计的人类早期第二语言学习认知过程的数据收集系统框架,并结合心理学、语言学的已有研究工作,针对中文母语学龄前儿童的英语第二语言学习设计了完整的数据收集系统,现已投入实际的数据收集。



图 2 首页页面

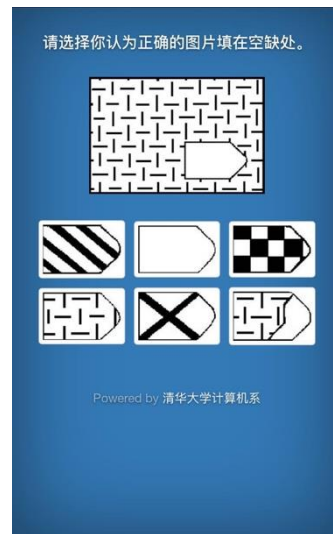


图 3 词汇评测页面

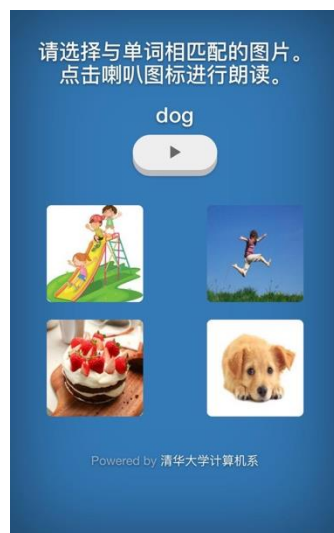


图 4 瑞文推理测验页面



图 5 工作记忆测验页面

游戏性设计使得用户从游戏中反馈中能够了解孩子的英语词汇发展水平相较于母语学龄前儿童学习的状况,同时可以轻松地从中收集到大规模数据用于相关研究。

值得强调的是,这一框架可以实现低成本的在线大规模数据收集,不受地域限制且不涉及个人隐私。收集过程中可以充分考虑将非语言智力、语言学习能力等变量融入到交互过程中,给用户的反馈内容也将包括对语言的学习指导,可以更好地实现双赢。此外,由于框架并不局限于某种语言或者场景的研究,该框架对于相关的实验设计有较好的借鉴意义;

未来我们将公布收集到的数据集,并且基于此数据集将初步的研究结果公布出来,用于促进语言学习认知方面研究的发展和进步。

## 参考文献

- [1] Łuniewska M, Haman E, Armon-Lotem S, et al. Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words?[J]. Behavior research methods, 2016, 48 (3): 1154-1177.

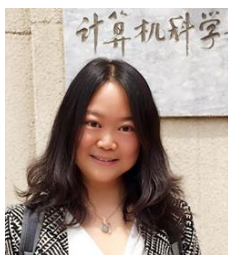
- [2] Braginsky M, Yurovsky D, Marchman V A, et al. Developmental Changes in the Relationship Between Grammar and the Lexicon[C]//CogSci. 2015, 256-261.
- [3] Schneider R, Yurovsky D, Frank M. Large-scale investigations of variability in children's first words[C]//CogSci. 2015, 2110-2115.
- [4] Fenson L, Bates E, Dale P S, et al. The MacArthur communicative development inventories[M]. Singular publishing Group, Incorporated, 1996.
- [5] Hills T T, Maouene M, Maouene J, et al. Longitudinal Analysis of Early Semantic Networks Preferential Attachment or Preferential Acquisition?[J]. Psychological Science, 2009, 20(6):729-739.
- [6] Beckage N M, Colunga E. Language Networks as Models of Cognition: Understanding Cognition through Language[M]// Towards a Theoretical Framework for Analyzing Complex Linguistic Networks. Springer Berlin Heidelberg, 2016:3-28.
- [7] Nelson D L, Mcevoy C L, Schreiber T A. The University of South Florida word association[J]. Alternative Higher Education, 1999, 3(3):154-160.
- [8] Macwhinney B. Second Language Acquisition and the Competition Model[J]. Tutorials in Bilingualism Psychological Perspectives, 1997, 21, 113-142.
- [9] 文敏琳. 论儿童母语习得与外语学习的异同[J]. 四川文理学院学报, 2012, (03):151-153.
- [10] 彭坚. 二语习得“关键期假说”的研究对儿童外语学习的启示[D]. 湖南: 湖南师范大学, 2010.
- [11] Von A L, Maurer B, Mcmillen C, et al. reCAPTCHA: human-based character recognition via Web security measures. [J]. Science, 2008, 321(5895):1465-1468.
- [12] Von Ahn L, Dabbish L. Designing games with a purpose[J]. Communications of the Acm, 2008, 51(8):58-67.
- [13] Von Ahn L, Ginosar S, Kedia M, et al. Improving Image Search with PHETCH[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE Xplore, 2007:IV-1209 - IV-1212.
- [14] Ahn L V, Liu R, Blum M. Peekaboom:a game for locating objects in images[C]// Sigchi Conference on Human Factors in Computing Systems. ACM, 2006:55-64.
- [15] Raven J C. Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive[J]. Unpublished master's thesis, University of London, 1936.
- [16] Miyake A, Shah P. Models of working memory: Mechanisms of active maintenance and executive control[M]. Cambridge University Press, 1999.
- [17] Kaplan R M, Saccuzzo D P. Standardized tests in education, civil service, and the military[J]. Psychological testing: Principles, applications, and issues 2009, 7: 325-327.
- [18] Ellis N C. Working memory in the acquisition of vocabulary and syntax: Putting language in good order[J]. The Quarterly Journal of Experimental Psychology: Section A, 1996, 49(1): 234-250.

## 作者简介:



**马为之** (1992—), 男, 清华大学博士生, 主要研究领域为用户画像、个性化推荐。Email: [mawz14@mails.tsinghua.edu.cn](mailto:mawz14@mails.tsinghua.edu.cn);

**Weizhi Ma**, born in 1992, Ph.D. candidate. His research interests include User Profiling and Personalized Recommendation. Email: [mawz14@mails.tsinghua.edu.cn](mailto:mawz14@mails.tsinghua.edu.cn)



**张敏 (通讯作者)** (1977—), 女, 清华大学副教授, 主要研究领域为信息检索与挖掘、个性化推荐与用户建模。Email: [z-m@tsinghua.edu.cn](mailto:z-m@tsinghua.edu.cn);

**Min Zhang (Corresponding author)**, born in 1977, Ph.D., Associate Professor. Her research interests include Web Information retrieval and Data mining, Personalized recommendation and User modeling. Email: [z-m@tsinghua.edu.cn](mailto:z-m@tsinghua.edu.cn)