

基于动画片语料的儿童词汇研究 Children's Chinese Vocabulary Study Based on a Corpus of Animated Films

许源海

NATTANON TEERAPANYAWATT

北京语言大学、国际中文教育研究院，中国

College of International Chinese Studies, Beijing Language and Culture University, P.R.China

E-mail: nattanon.nnt@gmail.com

Received: 03 October 2025 / Revised: 19 November 2025 / Accepted: 02 December 2025

摘要

本研究的目的在于（1）系统整理中国大陆地区播出的动画作品，构建中国动画片专用语料库；（2）对儿童动画片的用词情况进行深入分析；（3）建立中国动画核心词表。通过收集《大头儿子和小头爸爸》《喜羊羊与灰太狼》《熊出没》三部动画作品共计 30 季、2332 集语料，运用语料库语言学方法构建中国儿童动画专题语料库，并基于此开展儿童汉语词汇特征研究。语料分析表明，在汉字使用层面，高频字呈现显著集中趋势：使用频率最高的前五个汉字依次为“我”“不”“了”“的”“你”。高频字前 100 字与前 200 字的分布特征显示，均以一级汉字为主，其次是二级汉字。在词汇使用层面，高频词前五位均为单字词，具体为“我”“的”“了”“啊”“你”。对最高频 200 词的统计分析进一步证实，高频词主要一级词汇和二级词汇为主。三部作品的独特词汇分析发现《大头儿子和小头爸爸》的独有词汇主要集中于日常生活场景及家庭成员称谓；《喜羊羊与灰太狼》的独有词汇体系围绕“难题/挑战-伙伴协作-解决方案”；《熊出没》的独有词汇主要围绕野生动物与森林生态相关词汇。

关键词：动画片；语料库；儿童；词汇

ABSTRACT

The purposes of this study are to (1) systematically catalogue animated works broadcast in mainland China and construct a dedicated corpus of Chinese animation; (2) conduct an in-depth analysis of lexical usage in children's animated programs; and (3) establish a core vocabulary list for Chinese animation. Drawing on a dataset comprising 30 seasons and 2,332 episodes from Big Head Son and Small Head Dad, Pleasant Goat and Big Big Wolf, and Boonie Bears, the study employs corpus-linguistic methods to build a specialized corpus of Chinese children's animation and, on this basis, examines the lexical features of children's Chinese.

Corpus analysis reveals a pronounced concentration in the use of high-frequency characters: the five most frequent characters are “我”, “不”, “了”, “的”, and “你”. The distribution of the top 100 and top 200 high-frequency characters indicates that they are predominantly first-level characters, followed by second-level characters. At the lexical level, the five most frequent words are all monosyllabic items—specifically “我”, “的”, “了”, “啊” and “你”. Statistical analysis of the top 200 high-frequency words further confirms that they consist mainly of first-level and second-level vocabulary items.

An examination of the distinctive lexical items across the three animated series shows that Big Head Son and Small Head Dad features unique vocabulary centered on everyday life scenarios and kinship terms; Pleasant Goat and Big Big Wolf exhibits a distinctive lexical system revolving around “problems/challenges—peer collaboration—solutions”; and Boonie Bears contains unique vocabulary primarily associated with wildlife and forest ecology.

Keywords: animation, corpus linguistics, children’s vocabulary

1. 引言

国际中文教育在全球范围内迅速发展，泰国政府对此高度重视，已正式将中文教育纳入国民教育体系。汉语课程设置涵盖从幼儿园至大学的完整学段，导致教学对象呈现显著异质性，覆盖儿童至成人全年龄段。汉语教学需依据不同学习者群体的认知特征设计差异化教学方案与内容，尤其针对学龄前及小学低年级学习者，亟须开发适配其认知发展水平的专用教材与教学体系。近年来泰国低龄汉语学习者规模持续扩大，面向该群体的汉语教材出版量同步显著增长，泰国本土出版社与中国大陆出版社均积极投入针对性教材研发与出版工作。

但是，国际中文教育对教材和教学内容集中在成年人，针对儿童汉语教材和教学内容相关研究数量并不多，尤其是对儿童汉语词汇大纲研究数量并不多。国家汉办（2009:2）在《新中小学生汉语考试大纲》提出中小学生汉语考试分别为四级，一级水平应该掌握 80 个词汇、二级水平应该掌握 150 个词汇、三级水平应该掌握 300 个词汇、四级水平应该掌握 600 个词汇。崔永华（2014:50）提出，小学汉语教材词汇大纲的制定应遵循以下流程：明确教学时长、设定词汇总量、实施词汇分级与分类、编制核心词表等环节。刘华和郑婷（2017:70）以 12 套代表性的东南亚少儿华语教材为语料建设少儿华语教学主题分类词表。少儿华语主题词表中的一级指标包含个人信息、家庭、日常生活、学校生活、文化、动物与植物、自然与环境、科学技术。CHIH-SHENG CHEN 和 WEI GU（2019:27）通过对印度尼西亚学龄前儿童汉语词汇的研究发现，部分词汇存在分类错误或混淆现象；现有词表主要涵盖名词，部分涉及形容词和动词，但其他人称代词等词类未被包含。该研究提出的印度尼西亚学龄前儿童汉语词表包含 25 个主题类别，例如身体感觉、食物、动物、生活物品等。谢婧怡等人（2022:106）针对第二语言儿童口语词表的研发提出了相关编写原则，包括：（一）通用词表应集中反映儿童语言发展的共性，并为本土化教学词表的制定提供参

考依据：（二）通用词表编制需围绕意义主题展开，结合词频与义频进行构建；（三）通用词表应按照阶梯性原则，关注并适应不同年龄阶段学习者的认知特征；（四）通用词表在本土化设计时，应充分考虑学习者母语及文化背景的差异性。

由此可见，针对儿童汉语词汇的研究数量较少，其中部分研究聚焦于词汇教学及课堂应用，而关于词汇大纲及词表的研究则更为稀缺。本研究计划通过搜集儿童动画片来构建专用语料库，并运用词汇分析来探究其用词特点，从而为儿童汉语词汇研究提供实证支持。

2. 研究目的

本研究的研究目的主要有以下三点：

- （1）系统整理中国大陆地区播出的动画作品，构建中国动画片专用语料库；
- （2）对儿童动画片的用词情况进行深入分析；
- （3）建立中国动画核心词表。

3. 研究方法及数据来源

本研究通过搜集三部共 30 季、总计 2332 集的儿童动画片，运用语料库研究方法，构建了中国动画片语料库，并在此基础上对儿童汉语词汇进行了深入分析，为了了解儿童动画片的用字词情况。

（一）语料来源

本研究以中国大陆地区出品的三部具有代表性的儿童动画片为分析对象，分别为《大头儿子和小头爸爸》《喜羊羊与灰太狼》及《熊出没》。作品的选择标准包括：在中国大陆地区拥有较高的社会评价与广泛影响力；其中《喜羊羊与灰太狼》和《熊出没》具备较长的播映周期及多样化的剧情类型，而《大头儿子和小头爸爸》作为家庭题材动画片，有助于丰富家庭相关词汇的研究内容。所采集的动画片覆盖共计 30 季、2332 集，其中包括《大头儿子和小头爸爸》156 集、《喜羊羊与灰太狼》976 集以及《熊出没》1200 集。样本首播时间跨度为 1995 年至 2023 年（见表 1）。

表格 1 语料来源

序号	动画片系列名	动画片名	集	首播时间
1	喜羊羊与灰太狼	大头儿子和小头爸爸	156	1995 年
2		喜羊羊与灰太狼	530	2005 年
3		喜羊羊与灰太狼之嘻哈闯世界	60	2016 年
4		喜羊羊与灰太狼之发明大作战	60	2017 年
5		喜羊羊与灰太狼之深海历险记	60	2017 年
6		喜羊羊与灰太狼之奇幻天空岛	60	2018 年
7		喜羊羊与灰太狼之跨时空救兵	60	2019 年
8		喜羊羊与灰太狼之羊村守护者	60	2019 年
9		喜羊羊与灰太狼之奇趣外星客	60	2020 年
10		喜羊羊与灰太狼之羊羊趣冒险	26	2020 年



序号	动画片系列名	动画片名	集	首播时间
11	熊出没	熊出没 TV 版	52	2012 年
12		熊出没 老版	104	2012 年
13		熊出没之环球大冒险 TV 版	52	2012 年
14		熊出没之环球大冒险 老版	104	2012 年
15		熊出没之丛林总动员 TV 版	89	2013 年
16		熊出没之丛林总动员 老版	104	2013 年
17		熊出没之春日对对碰	52	2014 年
18		熊出没之冬日乐翻天	52	2015 年
19		熊出没之秋日团团转	52	2016 年
20		熊出没之夏日连连看	52	2016 年
21		熊出没之探险日记 S01	52	2017 年
22		熊出没之熊熊乐园 S01	52	2017 年
23		熊出没之熊熊乐园 S02	52	2018 年
24		熊出没之探险日记 S02	52	2019 年
25		熊出没之熊熊乐园 S03	52	2019 年
26		熊出没之怪兽计划 S01	52	2020 年
27		熊出没之熊熊乐园 S04	26	2021 年
28		熊熊帮帮团	52	2021 年
29		熊出没之怪兽计划 S02	52	2022 年
30		出没之小小世界	45	2023 年

(二) 研究工具

本研究搜集数据为视频文件 (.MP4)，运用了“录音啦”软件人工智能视频转文字技术功能，将视频内容精准转化为文字，并紧接着借助结巴 (JIEBA) 分词工具进行细致的分词与停用词剔除，最终通过汉语助研软件对所得语料数据进行全面统计分析。

4. 研究结果

本研究通过搜集中国大陆制作的儿童动画片，对所收集的动画片进行视频转录为文字材料。视频转录文字环节，本研究严谨地采用了人工智能的精确转录模式，确保文字转换的准确性。转录完成后，儿童动画片的文字材料被保存为两份：一份经过结巴分词工具的分词和停用词处理，作为词汇分析的语料库；另一份仅经过结巴分词工具的分词处理，作为参考语料库。

在分词处理过程中，结巴分词工具依据研究者提供的词典进行分词，并根据停用词表进行停用词处理。本研究专为结巴分词工具定制了词典，该词典参考《国际中文教育中文水平等级标准》编制，同时，停用词表的制定也参考了动画片中的片头、片尾歌词以及卡通角色名称，以确保分析的准确性。对第一份分析语料进行停用词处理的主要目的是更精确地分析动画片中的用词情况，确保角色名称和片头片尾的歌词不会对词汇统计造成干扰。在制定停用词表时，本研究决定停用完整的歌词句子，诸如“踏上了未知的旅途，无论遭遇何种艰难险阻，一同闯入异国的迷雾。”停用词表

中将包含“踏上了未知的旅途”“不管什么危难险阻”“一起闯入异国的迷雾”等句子作为停用词。在结巴（JIEBA）分词工具进行分词时，将依据提供的停用词表进行停用处理，若歌词与停用词表中的词汇不完全匹配，则不会进行停用，以避免过度停用词汇影响分词的准确性。

（一）儿童动画片词汇分析

1. 用字分析

本研究通过对动画片语料库内所有语料的分析，探讨了动画片的汉字使用情况使用汉语助研软件进行分析数据，发现本语料库 2,332 条数据包含了 3,671 汉字。为了筛选使用范围小以及在视频转换过程中可能出现的识别错误，本研究将删除在五条数据（集）或以下出现的汉字。经过筛选后剩下 2,728 汉字。按照汉字的使用频率前五汉字分别是：“我”在 2,168 条数据（集）被使用 61,368 次；“不”在 2,327 条数据（集）被使用 46,808 次；“了”在 2,279 条数据（集）被使用 46,795 次；“的”在 1,571 条数据（集）被使用 44,985 次；“你”在 1,945 条数据（集）被使用 38,661 次（见表 2）。

表格 2 儿童动画片字表前 50 字

汉字	频率	频次	文档数	文档频率	频序	汉字	频率	频次	文档数	文档频率	频序
我	0.029605701	61368	2168	0.929674099	1	大	0.006754488	14001	2202	0.944253859	26
不	0.022581535	46808	2327	0.997855918	2	去	0.006731814	13954	2163	0.927530017	27
了	0.022575264	46795	2279	0.977272727	3	小	0.006144698	12737	2165	0.92838765	28
的	0.021702067	44985	1571	0.673670669	4	快	0.00613119	12709	2235	0.958404803	29
你	0.018651186	38661	1945	0.834048027	5	子	0.005683013	11780	2088	0.895368782	30
啊	0.018586059	38526	1649	0.707118353	6	没	0.005578809	11564	2221	0.952401372	31
是	0.017027331	35295	2303	0.987564322	7	到	0.005559511	11524	2168	0.929674099	32
这	0.014250952	29540	2324	0.996569468	8	诶	0.005333735	11056	2112	0.905660377	33
们	0.013020758	26990	2201	0.943825043	9	吧	0.005270537	10925	1496	0.641509434	34
一	0.012744809	26418	2323	0.996140652	10	什	0.005124361	10622	1359	0.582761578	35
呀	0.012641569	26204	2240	0.960548885	11	儿	0.004819948	9991	2190	0.939108062	36
来	0.012343428	25586	2319	0.994425386	12	天	0.004770258	9888	2026	0.868782161	37
好	0.011719647	24293	2320	0.994854202	13	都	0.00475289	9852	1415	0.6067753	38
哎	0.011148934	23110	2300	0.986277873	14	羊	0.004637107	9612	1053	0.451543739	39
么	0.010873949	22540	1720	0.737564322	15	还	0.004488037	9303	1933	0.82890223	40
有	0.010077459	20889	2210	0.947684391	16	下	0.004438347	9200	2199	0.94296741	41
嗯	0.009180141	19029	1285	0.55102916	17	可	0.004426768	9176	1880	0.806174957	42
会	0.008049327	16685	2211	0.948113208	18	那	0.004382867	9085	1979	0.848627787	43
个	0.007651323	15860	2306	0.988850772	19	熊	0.004255988	8822	1186	0.508576329	44
就	0.007619966	15795	1870	0.801886792	20	上	0.004216429	8740	2169	0.930102916	45
要	0.007089776	14696	2093	0.897512864	21	太	0.00412911	8559	2151	0.92238422	46
俺	0.007040086	14593	1232	0.528301887	22	他	0.004083279	8464	1415	0.6067753	47
呃	0.006971099	14450	2038	0.873927959	23	怎	0.00397618	8242	1531	0.65651801	48
在	0.006908865	14321	1752	0.751286449	24	想	0.003878247	8039	2065	0.885506003	49
看	0.006871236	14243	2229	0.955831904	25	真	0.00387294	8028	2169	0.930102916	50



本研究在汉字统计的基础上对汉字进行等级标注，本次等级标注基于《国际中文教育中文水平等级标准》中的汉字表。根据《国际中文教育中文水平等级标准》（GF0025-2021），学习者中文水平被划分为三等九级，一级汉字新增 300 个汉字，二级汉字新增 300 个汉字累计 600 个汉字，三级汉字新增 300 个汉字累计 900 个汉字，四级汉字新增 300 个汉字累计 1200 个汉字，五级汉字新增 300 个汉字累计 1500 个汉字，六级汉字新增 300 个汉字累计 1800 个汉字，七至九级汉字新增 1,200 个汉字累计 3,000 个汉字。本研究主要通过两组数据对汉字等级进行分析：第一组数据基于儿童动画片，揭示了汉字表内各等级汉字的数量和比例；第二组数据则对比了儿童动画片汉字表与《国际中文教育中文水平等级标准》中的汉字表，以分析两者的比例关系。

通过对儿童动画片汉字表进行等级标注并分析，我们发现七至九级汉字数量最多，占儿童动画片字表的 27.27%，其次是一级汉字占 11.00% 和二级汉字占 10.89%。

另外，本研究通过对比分析儿童动画片汉字表与《国际中文教育中文水平等级标准》中的汉字表，发现《国际中文教育中文水平等级标准》中的一级汉字全部出现在儿童动画片字表中，二级汉字有 297 个出现在儿童动画片字表中，占《国际中文教育中文水平等级标准》二级字表的 99.00%，三级汉字有 294 个出现在儿童动画片字表中，占《国际中文教育中文水平等级标准》三级字表的 98.00%（见表 3）。

表格 3 儿童动画片汉字等级标注统计

汉字等级		汉字等级		汉字在《国际中文教育中文水平等级标准》字表中的占比	
		字	百分比	字汇量	百分比
初等	一级汉字	300	11.00	300	100.00
	二级汉字	297	10.89	300	99.00
	三级汉字	294	10.78	300	98.00
中等	四级汉字	292	10.70	300	97.33
	五级汉字	287	10.52	300	95.67
	六级汉字	280	10.26	300	93.33
高等	七—九级汉字	744	27.27	1,200	62.00
超纲		234	8.58		
合计		2,728	100.00		

本研究认为，仅通过分析字表整体的等级情况无法全面概括儿童动画片中汉字的使用情况，因为字表等级并未涵盖汉字在实际使用中的具体情况。本研究将截取使用频率最高的前 100 字和 200 字再进行统计汉字等级。分析结果显示，儿童动画片中，使用频率最高的前 100 字与前 200 字的汉字等级分布极为相似，均以一级汉字为主，其次是二级汉字，超纲汉字在前 100 字中较多，而三级汉字则在前 200 字中逐渐增多（见表 4）。通过统计可以发现儿童动画片的常用汉字基本在初

级汉字内。在使用频率最高的前 200 字出现的超纲汉字主要是：嗯（171）、俺（22）、呃（23）、诶（33）、啥（120）、咋（124）、哟（144）、哒（183）。

表格 4 儿童动画片前 100、200 汉字等级标注统计

汉字等级		前 100 字		前 200 字	
		字	百分比	字	百分比
初等	一级汉字	74	74.00	125	62.50
	二级汉字	15	15.00	37	18.50
	三级汉字	3	3.00	16	8.00
	四级汉字	1	1.00	5	2.50
中等	五级汉字	1	1.00	2	1.00
	六级汉字	1	1.00	2	1.00
高等	七-九级汉字	1	1.00	5	2.50
超纲字		4	4.00	8	4.00
合计		100	100.00	200	100.00

通过分析用字情况发现根据《国际中文教育中文水平等级标准》的分级，本次收集的语料用字范围覆盖了大部分汉字，其中初等和中等汉字的 90% 以及超过一半的高等汉字都包含在该标准的字表中，而超出标准范围的汉字仅占总汉字的 8.58%。从总体来看本次对动画片使用汉字分析可以发现动画片的用字情况基本在《国际中文教育中文水平等级标准》的分围内。进一步分析使用频率高的汉字发现，动画片常用的汉字基本集中在初等汉字，中高等汉字占比非常少。从使用频率较高的汉字来看，本次收集的三部动画片所用汉字难度较低，适合儿童或初级汉语学习者使用，完成一级汉字学习后，即可基本掌握动画片中的常用汉字。

2. 用词分析

本研究利用汉语助研软件，对动画片语料库内的全部语料进行了词汇使用情况的分析，结果显示，该语料库包含 2,332 条数据，总计 1,326,060 词次，涉及 51,631 个词汇。为了确保词汇表中的词汇具有足够的代表性，本研究删除了使用范围等于或小于 5 条语料（集）的词汇，经过严格筛选，我们最终保留了 10,294 个具有代表性的词汇。然而，在人工核实词汇表内容时，我们发现分词工具的表现不尽如人意，词表中包含了一些如“不会”（即“不”与“会”的组合）的组合词汇，该词汇《国际中文教育中文水平等级标准》和《现代汉语词典》没有收入。本研究将对词表进行等级标注，对于超出范围的词汇，我们将参照《现代汉语词典》进行收录情况的核实，若这些词汇未收录于《国际中文教育中文水平等级标准》或《现代汉语词典》中，则将被剔除。通过核等级标注和核对《现代汉语词典》词汇删除 4,495 条词汇，最终剩下 5,799 条词汇。词汇表的常用前 10 词如下：“我”“的”“了”“啊”“你”“嗯”“是”“呃”“哎呀”“俺”。（见表 5）

¹ 括号的数字是该汉字在字表的频序



表格 5 儿童动画片词表前 50 词

词语	频率	频次	文档数	文档频率	位序	词语	频率	频次	文档数	文档频率	位序
我	0.043103	44217	1287	0.551887	1	这个	0.005919	6072	1980	0.849057	26
的	0.040558	41606	1287	0.551887	2	把	0.005733	5881	1239	0.531304	27
了	0.03789	38869	1287	0.551887	3	不	0.005703	5850	1236	0.530017	28
啊	0.033495	34360	1287	0.551887	4	要	0.005628	5773	1241	0.532161	29
你	0.028397	29131	1286	0.551458	5	会	0.005495	5637	1165	0.499571	30
嗯	0.01855	19029	1285	0.551029	6	怎么	0.005425	5565	1219	0.522727	31
是	0.015652	16056	1287	0.551887	7	吗	0.005322	5460	1214	0.520583	32
呃	0.014061	14424	2037	0.873499	8	他	0.004982	5111	1148	0.492281	33
哎呀	0.011566	11865	2108	0.903945	9	来	0.004802	4926	1224	0.524871	34
俺	0.010713	10990	1201	0.515009	10	快	0.004211	4320	1181	0.506432	35
我们	0.010534	10806	1161	0.497856	11	不是	0.004159	4266	1811	0.776587	36
吧	0.010004	10263	1279	0.548456	12	啦	0.004151	4258	1724	0.73928	37
在	0.009934	10191	1268	0.543739	13	那	0.004089	4195	1187	0.509005	38
都	0.009151	9387	1256	0.538593	14	还	0.003984	4087	1206	0.517153	39
就	0.008858	9087	1278	0.548027	15	给	0.003853	3953	1145	0.490995	40
呀	0.008583	8805	1269	0.544168	16	看	0.003684	3779	1202	0.515437	41
有	0.00828	8494	1267	0.54331	17	吃	0.003597	3690	890	0.381647	42
什么	0.008038	8246	1252	0.536878	18	哦	0.003404	3492	1106	0.474271	43
哎	0.007954	8160	2040	0.874786	19	哼	0.003327	3413	1569	0.672813	44
去	0.007365	7555	1271	0.545026	20	说	0.003206	3289	1062	0.455403	45
你们	0.00705	7232	1209	0.518439	21	让	0.003135	3216	1115	0.47813	46
呢	0.006654	6826	1268	0.543739	22	羊	0.002917	2992	822	0.352487	47
好	0.006489	6657	1256	0.538593	23	啥	0.002889	2964	1007	0.431818	48
这	0.006256	6418	1247	0.534734	24	他们	0.002672	2741	843	0.361492	49
也	0.006126	6284	1258	0.539451	25	没有	0.002536	2602	1024	0.439108	50

通过标注《国际中文教育中文水平等级标准》的词汇等级和 YCT（少儿汉语考试）词汇等级，儿童动画片词汇表通过标注《国际中文教育中文水平等级标准》词汇等级发现超纲词汇数量最多高达 1,679 个词汇占总词汇的 28.95%，其次是七一九级词汇 1,059 个词汇占总词汇的 18.26% 和三级词汇 596 个词汇占总词汇的 10.28%。另外，标注 YCT 等级后，超纲词汇最多，达 5,363 个，占比 92.48%；其次是 YCT 四级词汇 211 个，占比 3.64%；YCT 三级词汇 107 个，占比 1.85%。（见表 6）

表格 6 儿童动画片词汇等级标注统计

等级	等级标准	YCT（少儿汉语考试）	
		词汇量	百分比
初等	一级词汇	422	7.28
	二级词汇	537	9.26
	三级词汇	596	10.28
中等	四级词汇	544	9.38
	五级词汇	509	8.78

等级	等级标准		YCT (少儿汉语考试)	
	词汇量	百分比	词汇量	百分比
高等	六级词汇	453	7.81	
	七-九级词汇	1,059	18.26	
	超纲词汇	1,679	28.95	5,363 92.48
	合计	5,799	100.00	5,799 100.00

本研究认为，仅通过分析词表整体的等级情况无法全面概括儿童动画片中词汇的使用情况，因为词表等级并未涵盖汉字在实际使用中的具体情况。本研究将截取使用频率最高的前 500 词再进行统计汉字等级。分析结果显示，儿童动画片词汇使用频率最高前 200 词，标注《国际中文教育中文水平等级标准》词汇等级后一级词汇数量最多有 96 个词汇占 48%，第二是二级词汇数量最多有 34 个词汇占 17.00%，第三是超纲词汇数量最多有 25 个词汇占 12.50%。通过标注 YCT 等级，词汇等级后超纲词汇数量最多有 110 个词汇占 55.00%，第二是 YCT 三级词汇数量最多有 28 个词汇占 14.00%，第三是 YCT 一级词汇数量最多有 26 个词汇占 13.00%（见表 7）。

表格 7 儿童动画片前 200 词汇等级标注统计

等级		等级标准		YCT（少儿汉语考试）	
		词汇量	百分比	词汇量	百分比
初等	一级词汇	96	48.00	26	13.00
	二级词汇	34	17.00	14	7.00
	三级词汇	16	8.00	28	14.00
	四级词汇	4	2.00	22	11.00
中等	五级词汇	8	4.00		
	六级词汇	7	3.50		
高等	七-九词汇	10	5.00		
	超纲词汇	25	12.50	110	55.00
	合计	200	100.00	200	100.00

通过建立儿童动画片词汇表可以发现，动画片使用的词汇量在五千词左右，词汇难度分布统计显示，以等级大纲为标注依据，词表难度分布较为均匀，初等、中等词汇和超纲词汇的平均占比均在 25% 以上，而高等词汇的比例约为 18%，比其他等级低约 5 个百分点。但使用 YCT 词汇进行标注发现本次得出的大纲基本超 YCT 大纲，在大纲内的词汇低于 8%。标注结果表明词汇表难度较高，但研究者认为，词汇的使用频率和使用范围也是影响难度统计的重要因素。本研究进一步统计高频词的难度，发现使用率前 200 词的词汇难度的 73.00% 是初等词汇，其次是超纲词汇占 12.50%，最后是中等和高等词汇。使用率前 500 词的词汇难度情况也相似初等词汇比例最多占 65.60%。从这两组数据来看儿童动画片词汇表高频词以初等词汇为主。

(二) 三部儿童动画片用词分析

用词分析主要为了了解该动画片的用词情况,通过单独建设三部儿童动画片的语料库,并进行建设词表,最后把三部动画片词表进行比较。建设和处理词汇表的步骤如下:第一步,基于动画片语料库,按语料来源分类,为每部动画片建立文件夹。第二步,使用汉语助研软件对每一部动画片进行用词分析,得出用词报告。第三步,删除用词报告中使用范围小于或等于5条语料的词汇。第四步,对词表进行等级标注,核对《现代汉语词典》删除未收录于《国际中文教育中文水平等级标准》及《现代汉语词典》的超纲词汇。通过建立三部动画片词汇表后《大头儿子和小头爸爸》词汇表包含757个词汇,其中一级词汇数量最多占总表的35.93%。《喜羊羊与灰太狼》词汇表包含3,734个词汇,其中七-九级词汇数量最多占总表的18.40%。《熊出没》词汇表包含3,078个词汇,其中超纲词汇数量最多占总表的27.03%(见表8)。

表格 8 三部儿童动画片词汇表等级标注统计

等级		《大头儿子和小头爸爸》		《喜羊羊与灰太狼》		《熊出没》	
		词汇量	百分比	词汇量	百分比	词汇量	百分比
初等	一级词汇	272	35.93	403	10.79	240	7.80
	二级词汇	201	26.55	507	13.58	184	5.98
	三级词汇	93	12.29	500	13.39	314	10.20
	四级词汇	48	6.34	407	10.90	272	8.84
中等	五级词汇	47	6.21	350	9.37	293	9.52
	六级词汇	19	2.51	271	7.26	254	8.25
高等	七-九级词汇	24	3.17	687	18.40	689	22.38
超纲词汇		53	7.00	609	16.31	832	27.03
合计		757	100	3,734	100	3,078	100

本研究进一步统计三部儿童动画片词汇表的使用率最高200次的词汇等级情况,发现三部儿童动画片一级词汇占比最高。另外根据统计《大头儿子和小头爸爸》《喜羊羊与灰太狼》这两部词汇表四级至九级词汇数量较少(见表9)。

表格 9 三部儿童动画片前200词汇等级标注统计

等级		《大头儿子和小头爸爸》		《喜羊羊与灰太狼》		《熊出没》	
		词汇量	百分比	词汇量	百分比	词汇量	百分比
初等	一级词汇	120	60.00	103	51.50	46	23.00
	二级词汇	39	19.50	42	21.00	9	4.50
	三级词汇	8	4.00	17	8.50	35	17.50
	四级词汇	3	1.50	2	1.00	15	7.50
中等	五级词汇	4	2.00	6	3.00	21	10.50
	六级词汇	3	1.50	6	3.00	11	5.50
高等	七—九级词汇	6	3.00	8	4.00	21	10.50
超纲词汇		17	8.50	16	8.00	42	21.00

等级	《大头儿子和小头爸爸》		《喜羊羊与灰太狼》		《熊出没》	
	词汇量	百分比	词汇量	百分比	词汇量	百分比
合计	200	100	200	100.00	200	100

为了更了解三部儿童动画片用词情况，将把三个词表进行交叉标注找出三部动画片的独有词汇。标注后发现《大头儿子和小头爸爸》共有 19 个独有词汇，例如“转眼”“围裙”“阁楼”“报纸”“上班”。《喜羊羊与灰太狼》共有 1,564 个独有词汇，例如“难题”“难”“时空”“家人”“散”。《熊出没》共有 1,050 个独有词汇，例如“狗熊”“熊熊”“光头”“发财”“成就”。（见表 10）

表格 10 三部儿童动画片词表独有词汇频率前 20

序号	《大头儿子和小头爸爸》	《喜羊羊与灰太狼》	《熊出没》
1	转眼	难题	狗熊
2	围裙	难	熊熊
3	阁楼	时空	光头
4	报纸	家人	发财
5	上班	散	成就
6	伯伯	摔倒	团子
7	邮递员	本	锯子
8	请进	将军	伐木
9	冰激凌	柔软	陪伴
10	祝贺	面前	猫头鹰
11	早晨	慌乱	松鼠
12	向上	与	黑熊
13	里头	难以想象	人参
14	儿童	草	星辰
15	商店	挑战	丛林
16	早饭	努力	人类
17	窗帘	伙伴	平淡
18	全家	创造	姑娘
19	母子	班长	王冠
20	-	而且	硬币

通过分析《大头儿子和小头爸爸》的用词情况发现，通过统计和人工处理共有 757 个词汇，词汇难度分布在初等词汇和中等词汇。通过对比发现《大头儿子和小头爸爸》的词汇量少于另外两部动画片，在词汇难度方面，高等词汇和超纲词汇相对较少，而初等词汇占据主导地位。高频词汇统计也发现前 200 词也集中在初等词汇，中等和高等词汇数量较少。经过三部互相标注发现，《大头儿子和小头爸爸》独有词汇数量较少只有 19 个生词。从独有词汇可以发现该动画片内容围绕日常生活和家庭成员，例如“围裙”“上班”“早饭”“伯伯”。

通过分析《喜羊羊与灰太狼》的用词情况，通过统计和人工处理共有 3,734 个词汇，进一步分析词汇难度发现该动画片的初等和中等词汇较多在一千词左右，高等词和超纲词在六百词左右。从用词难度可以发现《喜羊羊与灰太狼》的整体词汇难度会高于《大头儿子和小头爸爸》，《喜羊羊与灰太狼》的整体难度在二级到三级。高频词汇统计发现，该动画片前 200 个高频词主要集中在一级和二级词汇。经过三部互相标注发现，《喜羊羊与灰太狼》独有词汇数量最多一个 1,564 个词汇，独有词汇占整个词汇表的二分之一。通过分析该动画片的高频独有词汇可以发现，该动画片主要围绕主角遇到“难题”或者“挑战”通过“伙伴”的“努力”才能决绝“难题”。

通关分析《熊出没》的用词情况，发现该动画片比前两部动画片难。整体词汇难度分布均匀，初等、中等、高等和超纲词汇基本在 20% 左右，超纲词汇占比最高，占词表的 27.03%，其次是中等词汇占 26.61%。高频词汇统计也发现，该动画片前 200 个高频词难度比前两部动画片难，初等词汇只占 45.00% 前两部动画片的初等词汇均在 80.00% 以上。经过三部互相标注发现，《熊出没》的独有词汇占整个词汇表的三分之一。通过分析该动画片的高频独有词汇可以发现，该动画片的高频独有词汇围绕着野生动物和森林相关词汇，例如“狗熊”“猫头鹰”“松鼠”“黑熊”“伐木”“人参”“丛林”“山坡”等。

5. 总结

本研究通过系统收集中国大陆动画片的相关评价及数据，遴选在中国大陆具有较高社会评价和广泛影响力的代表性作品，分别是《大头儿子和小头爸爸》《喜羊羊与灰太狼》以及《熊出没》。研究对象覆盖上述三部动画共计 30 季、2332 集。所有视频资料均经过转写为文本，并依次完成分词与停用词处理，最终建立起中国动画片专用语料库。

基于中国动画片专用语料库，对儿童动画片用字词分析和分析各个动画片的用词情况进行。分析动画片用字情况发现，使用频率最高的前五个汉字主要是‘我’‘不’‘了’‘的’‘你’。语料库内包含《国际中文教育中文水平等级标准》的一级至六级的 90% 以上，七至九级的 60%。通过进一步分析发现，高频字前一百字和前两百字主要以一级汉字和二级汉字为主。在分析儿童动画片的用词发现高频词前 5 词汇均为单字词汇分别是“我”“的”“了”“啊”“你”。通过对词表进行难度标注，参照《国际中文教育中文水平等级标准》和 YCT 等级，发现词表中超纲词汇数量最多，其次是七至九级词汇。通过标准 YCT 等级也发现超纲词汇数量最多。进一步统计使用频率为最高 200 词发现，高频词主要以《国际中文教育中文水平等级标准》一级词汇和二级词汇为主，YCT 词汇以超纲词汇为主。独有词汇分析发现《大头儿子和小头爸爸》独有词汇围绕日常生活和家庭成员，例如“围裙”“上班”“早饭”“伯伯”。《喜羊羊与灰太狼》独有词汇主要围绕主角遇到“难题”或者“挑战”通过“伙伴”的“努力”才能决绝“难题”。《熊出没》有词汇围绕着野生动物和森林相关词汇，例如“狗熊”“猫头鹰”“松鼠”“黑熊”“伐木”。

基于中国动画片专用语料库用词分析，本研究建立《中国动画核心词表》，该词表包含《大头儿子和小头爸爸》、《喜羊羊与灰太狼》以及《熊出没》使用频率最高前 500 词，按照词汇难度进行排列

一级词汇: 我、的、了、你、是、我们、吧、在、都、就、有、什么、去、你们、呢、好、这、也、不、要、会、怎么、吗、他、来、快、那、还、给、看、吃、说、他们、没有、想、现在、走、知道、没、和、回来、找、做、这里、能、再、谁、到、真的、跑、着、对、时间、小、很、东西、还是、爸爸、等、帮、别、太、大、跟、今天、出来、哪儿、还有、上、起来、爱、人、叫、多、用、们、一样、天气、真、先、找到、最、谢谢、拿、个、这儿、时候、这些、里、天、儿子、不用、对不起、打、准备、妈妈、听、干什么、请、好吃、回去、喜欢、是不是、事、马上、难、放、再见、过、哥哥、朋友、地方、号、出去、帮忙、孩子、看到、一、笑、哪里、送、告诉、爷爷、先生、睡觉、回家、您、买、明天、睡、家、家人、本、下、站、看见、点、干、比、从、觉得、哪、那边、那些、饿、新、飞、树、一会儿、住、前、进去、中、休息、不对、行、一点儿、小朋友、玩儿、最后、手、出、开、爸。

二级词汇: 啊、让、一定、这么、可以、又、大家、这样、不要、因为、永远、办法、更、可是、开心、它、所有、才、得、已经、原来、那么、自己、像、怎么办、太阳、为什么、虽然、不行、只、只要、好像、应该、过来、刚才、怎么样、难题、可能、以后、小心、不过、所以、如果、的话、心情、发现、该、问题、加油、完、为、带、变成、喂、放心、怕、当、掉、漂亮、快乐、相信、但是、不错、站住、样子、礼物、面前、舒服、出发、脸、一直、事情、声音、就要、送给、想到、但、机会、过去、草、安全、完成、然后、算、不如、以为、离开、弄、努力、疼、而且。

三级词汇: 把、羊、就是、力量、被、抓、救、难道、危险、只是、游戏、开始、终于、发明、追、下来、到底、里面、当然、好好、只有、美好、成功、赶快、表演、要是、世界、为了、才能、玩具、石头、赶紧、苹果、害怕、下去、跳、外面、容易、木头、其实、敢、奇怪、死、没用、欢乐、愿望、突然、抓住、绝对、种子、破坏、连、武器、需要、发生、继续、拍、各位、哈哈、比赛、上面、香蕉、香、美丽、后面、精彩、果然、心、机器、保护、痛、创造、反正。

四级词汇: 呀、老婆、森林、怪、没错、勇敢、看来、宝宝、帽子、却、散、阻止、停下、老公、宝贝、竟然、叔叔、打败、亲爱、兄弟、快递、误会、笨、既然、担心、之前、说不定、大哥、挑战、失败、伙伴、擦、垃圾、力气、梦。

五级词汇: 熊、臭、欣赏、居然、厉害、可、聪明、拥有、糟、肯定、天才、讨厌、骗、神奇、再也、不许、躲、能量、摔倒、蛋糕、扔、棒、枪、逃、偷、总算、收拾、机器人、摘、耳朵、咬、将、搞、瞧。

六级词汇: 啦、就算、救命、笑容、嘛、王子、国王、哇、前方、不见、有没有、懒、捉、藏、欺负、算了、将军、与、捡、家园、挖。

七至九级词汇: 哎呀、哎、哦、哼、狼、飞翔、可恶、沮丧、蜂蜜、嘿、萝卜、砍、笨蛋、时空、冒险、家伙、放过、无敌、柔软、慌乱、而已、惩罚、难以想象、天地、乐园、怪物、青蛙、蘑菇、美味、倒霉、捣乱、蜜蜂、竟敢、彩虹。

超纲词汇: 嗯、呃、俺、这个、不是、啥、咋、一下、哎哟、哒、狗熊、真是、玩、没事、熊熊、大王、那个、追捕、一起、嘿嘿、呦、呐、喽、不了、耶、哈、不想、儿、呦、哟、咦、提取、奔放、绵羊、乖乖、不在、老虎、一天、嗨、开门、蝴蝶、猴子、果子、光头、修好、山羊、跟着。

6. 讨论

通过对儿童动画片语料的系统分析, 研究结果显示, 高频词汇均归属于初级词汇范畴, 并符合中小学各阶段的难度标准。因此, 中小学教师可据此参考《核心词汇表》进行教学内容设计。在该词表中出现的部分超纲词语主要包括感叹词(如“嗯”、“呃”、“啥”、“咋”、“哎哟”等), 因未被《国际中文教育中文水平等级标准》及 YCT 等级大纲收录, 但鉴于其语用难度较低, 教师和教材编写者可以适当补充, 并结合语音教学帮助学生习得词义。另一类超纲词汇涉及动物类词汇(如“狗熊”、“绵羊”、“老虎”、“蝴蝶”、“猴子”等), 同样未被前述标准和大纲涵盖。考虑到相关词汇认知负担较大, 建议在教学过程中配合图片等多模态手段, 以降低理解和记忆难度。

依据本研究对三部动画片词汇难度的系统分析, 汉语教师在教学资源选择时可据学习者汉语水平作出恰当判断。对于初学者或零基础学习者, 推荐优先选用《大头儿子和小头爸爸》; 汉语能力处于 HSK1 至 3 级的学习者则可选用《大头儿子和小头爸爸》或《喜羊羊与灰太狼》; 若学习者已达 HSK3 级及以上, 则建议考虑《喜羊羊与灰太狼》和《熊出没》。在具体择材过程中, 需综合评估词汇难度、学生年龄特征以及兴趣偏好等因素。如所选资源词汇难度高于学生实际水平, 教师在观看动画片前针对高频关键词进行预讲解, 以提升语言习得效果。此外, 根据 Mayer (2009) 提出的多媒体学习理论, 每一信息通道(听觉与视觉)的容量有限, 而双通道并行能够提升学习成效。因此, 建议教师在组织观看动画片时, 引导学习者同步阅读字幕, 以促进其对内容的理解与掌握。

然而, 本研究存在以下局限: 首先, 需结合中小学汉语课程的实际教学时长进一步核定词汇量基准; 其次, 应对词汇进行系统分类以便教材编写者及教师参考。

此外, 本研究通过对三部儿童动画片的词汇特征分析, 提出各片的词汇难度评估表。汉语教师在开展视听课程教学时, 可依据学生的汉语水平选取对应难度的动画资源。研究结果表明, 学生观看符合其语言认知水平的动画片, 既能提升汉语学习兴趣, 亦可促进词汇习得与应用能力。

未来研究方向建议: 第一, 可以对动画片单集进行细粒度的词汇难度分级与关键词提取。鉴于本研究目前以“部”为单位进行词汇分析, 而实际教学环节中通常按“集”作为资源单位, 未来研究可选择某一部动画片作为语料, 对其每一集展开系统的词汇等级标注及关键词提取工作。

通过建立每集动画片的词汇表，教师能够根据对应词表开展针对性的词汇教学，并可将动画片中的对话内容选作课文材料使用；第二，可以制定针对中小学阶段的动画资源分级指南，为汉语教师提供科学的参考体系。后续研究可通过拓展语料来源，系统收集涵盖多种题材的儿童动画片，并对每一集的词汇难度与话题进行标注和统计，从而形成具有权威性的动画资源分级指南。教师可依据具体教学话题及词汇难度，有针对性地选择适宜的动画片资源。

参考文献

- 崔永华. (2014). 小学沉浸式汉语教材词汇大纲研制思路. [J]. 国际汉语教学研究. 2014, (02): 48-56.
- 国家汉办/孔子学院总部. (2009). 新中小学生汉语考试大纲. 商务印书馆.
- 刘华、郑婷. 少儿华语教学主题分类词表构建. [J]. 华文教学与研究. 2017, (01): 68-74.
- 谢婧怡、朱志平、齐元涛. (2022). 二语儿童口语词表研制的理念与原则. [J]. 语言文字应用. 2022, (03): 100-113.
- CHENCHIH-SHENG, & GUWEI. Exploring Mandarin Chinese Vocabulary that Children in Kindergarten Should Learn in Indonesia. [J]. US-China Foreign Language. 2019, 17(1): 20-29.
- MayerRichard E.. Multimedia Learning (second Edition). [M]. Cambridge: Cambridge University Press, 2009.

	Name and Surname (姓名): NATTANON TEERAPANYAWATT
	Highest Education (最高学历): Doctorial students
	University or Agency (任职院校或单位): College of International Chinese Studies, Beijing Language and Culture University, P.R. China
	Field of Expertise (专业领域): Teaching Chinese as a Second Language
	Address (地址): 15 Xueyuan Road, Haidian District, Beijing 100083

