K-5 Chinese **Word Frequency Dictionary** for L2 Learners

Yalan King, Mandarin Institute
Magaly Lavadenz, Loyola Marymount University
Ping Liu, California State University Long Beach
Learning to read is the central achievement of early elementary schooling, and then using reading to learn...
Do you recognize these students?
Differences in characteristics of word **comprehension** impacts reading ability and **comprehension**.

- *Charles Perfetti*
  *Universal Theory of Reading*

“*These reading glasses don’t work. I got a C in Reading Comprehension.*”
The Challenge

We want our students to choose the *right level* books to reinforce and challenge their reading but

- Not enough reading materials
- Not easy to find
- Not leveled by standards
- Not easy to match with students’ developmental reading level and interests
If we know how well a student can read (performance) and the vocabulary levels of specific books/texts, we can predict how well that student will comprehend the book/text.

- **Text Measure**
  Industry standard measurement on difficulty of book

- **Reader Measure**
  Industry standard assessment on student’s reading ability
Key Features

- A **common scale** for measuring literacy and literacy activities
- A **developmental scale** for maintaining longitudinal perspectives on growth
- Enables **targeted literacy activities** to enhance growth
- Promotes **personalized instruction**
Phase I
- Word frequency count for the character corpus from children’s reading

Phase II
- Develop/Administer/Score/Analyze Vocabulary and Reading Comprehension tests to word frequency bands

Phase III
- Correlate word frequency, sentence length, other variables, reading scores to determine Text Complexity

Phase IV
- Validate and test Text Complexity and Reading Levels

Phase V
- Develop online Chinese Readability Assessment System to be used by K-5 Chinese Teachers
Our Strategy: A Collaborative Pioneering Effort

Development, validation and testing of a tool to determine text complexity in grades K-5

- Phase I: Develop K-5 word frequency dictionary
  - Select children’s reading materials
  - Input text materials into digital format
  - Calculate word & character frequency
The Chinese Readability Scale Team

Helen Shen
Ping Liu
L2 Chinese Literacy Expert
Chinese Linguist, Dual Immersion Curriculum, Instruction & Assessment Expert

Magaly Lavadenz
Yalan King
Co-Director
Co-Director

Master Practitioners
Research Assistants

K-5 Chinese Word Frequency Dictionary for L2 Learners
Some Criteria for Leveling Books

- Length
- Layout
- Structure & Organization
- Illustrations
  - Words / Characters
  - Graded Vocabulary List
  - Phrases and Sentences
  - Literary Features
  - Content and Theme
Collaborate with Master Chinese Language Teachers

- Narrative vs. Informational
- Categories by level
  - i. Instructional
  - ii. Independent
  - iii. Frustration
- Instructional Category by grade
  - i. Alignment with Content Standards
  - ii. Reading Comprehension questions
  - iii. Vocabulary
- Building collection of material that align to levels

K-5 Chinese Word Frequency Dictionary for L2 Learners
Huge Challenge

- Found software that could be adapted to help us calculate word frequency
  - Not possible to scan text
  - Input must be in Word format
- 1 hour to input 1700 characters
  - ~1500 hours of input time
- Put out an emergency call and secured 8 Research Assistants who helped us to input selected books
Our Accomplishments

- Created template for K-5 Chinese Reading classifications

- Developed High Frequency Words Database

- Assessed Text Complexity of Chinese children's reading materials

- Categorized and Aligned books towards the definition of corpus

- Sample books for each classification

- Created a corpus of 2.6M characters across K-5 classifications

- Determination of High Frequency “bands” for K-5 vocabulary
Using the K-5 Chinese WFD

- Predict student’s reading comprehension by grade
- Measure text complexity – teachers can determine if a reading text is at a certain level
- Provide vocabulary guidelines for textbook writing and reading material selection
- Inform the development of assessments

**Online Searchable WFD**  www.MandarinInstitute.org
## Summary K-5 Chinese WFD Bands for L2 Learners

<table>
<thead>
<tr>
<th>GRADE</th>
<th>WORDS FROM LOWER GRADE(S)</th>
<th>NEW WORDS</th>
<th>TOTAL = WORDS FROM LOWER GRADE(S) + NEW WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>0</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>1</td>
<td>300</td>
<td>400</td>
<td>700</td>
</tr>
<tr>
<td>2</td>
<td>300+400=700</td>
<td>500</td>
<td>1200</td>
</tr>
<tr>
<td>3</td>
<td>300+400+500=1200</td>
<td>600</td>
<td>1800</td>
</tr>
<tr>
<td>4</td>
<td>300+400+500+600=1800</td>
<td>700</td>
<td>2500</td>
</tr>
<tr>
<td>5</td>
<td>300+400+500+600+700=2500</td>
<td>849</td>
<td>3349</td>
</tr>
</tbody>
</table>
### WFD Word List by Grade Level

<table>
<thead>
<tr>
<th>序号</th>
<th>词</th>
<th>R1</th>
<th>R2</th>
<th>拼音</th>
<th>R1</th>
<th>R2</th>
<th>出现次数</th>
<th>新增频次</th>
<th>总频次</th>
<th>频率</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>的</td>
<td>1</td>
<td>1</td>
<td>de</td>
<td>1</td>
<td>1</td>
<td>8220</td>
<td>8220</td>
<td>0.0032880</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>了</td>
<td>1</td>
<td>1</td>
<td>le</td>
<td>1</td>
<td>1</td>
<td>4878</td>
<td>4927</td>
<td>0.0019708</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>一</td>
<td>1</td>
<td>1</td>
<td>yī</td>
<td>1</td>
<td>1</td>
<td>3281</td>
<td>3449</td>
<td>0.0013796</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>我</td>
<td>1</td>
<td>1</td>
<td>wǒ</td>
<td>1</td>
<td>1</td>
<td>3261</td>
<td>3261</td>
<td>0.0013044</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>是</td>
<td>1</td>
<td>1</td>
<td>shì</td>
<td>1</td>
<td>1</td>
<td>2633</td>
<td>2685</td>
<td>0.0010740</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>在</td>
<td>1</td>
<td>1</td>
<td>zài</td>
<td>1</td>
<td>1</td>
<td>2174</td>
<td>2239</td>
<td>0.0008956</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>不</td>
<td>1</td>
<td>1</td>
<td>bù</td>
<td>1</td>
<td>1</td>
<td>2028</td>
<td>2055</td>
<td>0.0008220</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>说</td>
<td>1</td>
<td>1</td>
<td>shuō</td>
<td>1</td>
<td>1</td>
<td>1794</td>
<td>1817</td>
<td>0.0007268</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>个</td>
<td>1</td>
<td>1</td>
<td>gè</td>
<td>1</td>
<td>1</td>
<td>1769</td>
<td>1779</td>
<td>0.0007116</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>他</td>
<td>1</td>
<td>1</td>
<td>tā</td>
<td>1</td>
<td>1</td>
<td>1747</td>
<td>1769</td>
<td>0.0007076</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>到</td>
<td>1</td>
<td>1</td>
<td>dào</td>
<td>1</td>
<td>1</td>
<td>1513</td>
<td>1599</td>
<td>0.0006396</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>着</td>
<td>1</td>
<td>1</td>
<td>zhe</td>
<td>1</td>
<td>1</td>
<td>1468</td>
<td>1467</td>
<td>0.0005872</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>有</td>
<td>1</td>
<td>1</td>
<td>yǒu</td>
<td>1</td>
<td>1</td>
<td>1467</td>
<td>1467</td>
<td>0.0005868</td>
<td></td>
</tr>
</tbody>
</table>

Download word frequency list [www.MandarinInstitute.org](http://www.MandarinInstitute.org)
<table>
<thead>
<tr>
<th>SID</th>
<th>Word</th>
<th>Pinyin</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3349</td>
<td>楼</td>
<td>lóu</td>
<td>0.0000016</td>
</tr>
<tr>
<td>3348</td>
<td>只得</td>
<td>zhíděi</td>
<td>0.0000016</td>
</tr>
<tr>
<td>3347</td>
<td>自由自在</td>
<td>zìyóuzizài</td>
<td>0.0000016</td>
</tr>
<tr>
<td>3346</td>
<td>漫长</td>
<td>màncháng</td>
<td>0.0000016</td>
</tr>
<tr>
<td>3345</td>
<td>探</td>
<td>tàn</td>
<td>0.0000016</td>
</tr>
<tr>
<td>3344</td>
<td>得出</td>
<td>déchū</td>
<td>0.0000016</td>
</tr>
</tbody>
</table>

Displaying rows 1 to 10 of 3,349
Using the K-5 Chinese *Word Frequency Dictionary for L2 Learners*
Application

小蝌蚪找妈妈

K-5 Chinese Word Frequency Dictionary for L2 Learners
春天来了，青蛙妈妈睡醒了，跳进池塘，生下很多圆圆黑黑的卵。

后来，卵变成了小蝌蚪，他们看见小鸭子跟着鸭妈妈游水，也要去找妈妈。

鸭妈妈告诉小蝌蚪："你们的妈妈有两只大眼睛，嘴巴又宽又大。"

小蝌蚪高高兴兴地去找妈妈。

小蝌蚪看见红鲤鱼，直叫 "妈妈"，红鲤鱼却说："你们的妈妈有四条腿。"
小蝌蚪又叫乌龟“妈妈”，乌龟说：“你们的妈妈肚皮是白的。”
小蝌蚪又遇见大白鹅，直叫“妈妈”，大白鹅说：“你们的妈妈穿着绿衣服。”
一会儿，青蛙妈妈游来了，小蝌蚪甜甜的喊“妈妈”。
可小蝌蚪一看，妈妈有四条腿，觉得自己一点也不像妈妈，青蛙妈妈说：“你们还没长大。”
小蝌蚪跟着妈妈做游戏戏水，慢慢地也长出了四条腿，他们长大了。
Why & how to select vocabulary?

I. Connect to content standards
II. Apply Word Frequency Dictionary (WFD)
III. Prepare students to use words to express understanding or ideas/thoughts
IV. Include selected words in assessment
I. Connection to ACTFL Proficiency Guidelines 2012: Novice High

- Novice High speakers respond to simple, direct questions or requests for information. They are also able to ask a few formulaic questions.
- Novice High speakers are able to express personal meaning by relying heavily on learned phrases or recombinations of these and what they hear from their interlocutor. Their language consists primarily of short and sometimes incomplete sentences in the present, and may be hesitant or inaccurate.
- Novice High writers are able to recombine learned vocabulary and structures to create simple sentences on very familiar topics, but are not able to sustain sentence-level writing all the time.
- Readers at the Novice High sublevel are typically able to derive meaning from short, non-complex texts that convey basic information for which there is contextual or extralinguistic support.

K-5 Chinese Word Frequency Dictionary for L2 Learners
Connection to California Next Generation Science Standards: Grade 1

Crosscutting Concepts

**Patterns:** Patterns in the natural and human designed world can be observed, used to describe phenomena, and used as evidence. (1-LS1-2)

Science and Engineering Practices

**Scientific Knowledge is Based on Empirical Evidence**
Scientists look for patterns and order when making observations about the world. (1-LS1-2)
Connection to California Next Generation Science Standards: Grade 1

Crosscutting Concepts

**Patterns:** Patterns in the natural and human designed world can be observed, used to describe phenomena, and used as evidence. (1-LS1-2)

Science and Engineering Practices

**Scientific Knowledge is Based on Empirical Evidence**
Scientists look for patterns and order when making observations about the world. (1-LS1-2)
Connection to California Common Core State Standards
ELA/Literary: Grade 1

- **RI.1.1** Ask and answer questions about key details in a text. (1-LS1-2)
- **RI.1.2** Identify the main topic and retell key details of a text. (1-LS1-2)
- **RI.1.10** With prompting and support, read informational texts appropriately complex for grade. (1-LS1-2)
  a. Activate prior knowledge related to the information and events in a text. **CA**
  b. Confirm predictions about what will happen next in a text. **CA**
- **W.1.7** Participate in shared research and writing projects (e.g., explore a number of “how-to” books on a given topic and use them to write a sequence of instructions). (1-LS1-1)
Connection to California Common Core State Standards
ELA/Literary: Grade 1

- **RI.1.1** Ask and answer questions about key details in a text. (1-LS1-2)
- **RI.1.2** Identify the main topic and retell key details of a text. (1-LS1-2)
- **RI.1.10** With prompting and support, read informational texts appropriately complex for grade. (1-LS1-2)
  a. Activate prior knowledge related to the information and events in a text. **CA**
  b. Confirm predictions about what will happen next in a text. **CA**
- **W.1.7** Participate in shared research and writing projects (e.g., explore a number of “how-to” books on a given topic and use them to write a sequence of instructions). (1-LS1-1)
II: Apply WFD: Grade 1 (700 in total)

Word match in selected text:
14* 小，19 看，34 妈妈，75 水，82 找
96 看见，135 四，178 鱼，179 眼睛，271 黑
282 春天，337 红色，357 衣服，
418 嘴，467 青蛙，496 圆，505 绿色
611 游泳，622 鸭，625 鸭子，688 黑色

SID = word sequence number  (smaller number = higher frequency)

K-5 Chinese Word Frequency Dictionary for L2 Learners
http://mandarininstitute.org
Other related words:
195 动物，249 变成
353 终于，408 母亲
457 颜色，477 到处
486 不同，488 好多
610 尾巴，662 遇到，676 寻找
### WFD word search: First grade

| 456 | 恒星 | 1 | 1 | héngxīng | 1 | 1 | 44 |
| 457 | 颜色 | 1 | 1 | yánsè | 1 | 1 | 44 |
| 458 | 换 | 1 | 1 | huàn | 1 | 1 | 44 |
| 459 | 久 | 1 | 1 | jiǔ | 1 | 1 | 44 |
| 460 | 赶快 | 1 | 1 | gǎnkuài | 1 | 1 | 44 |
| 461 | 应该 | 1 | 1 | yīnggāi | 1 | 1 | 44 |
| 462 | 希望 | 1 | 1 | xīwàng | 1 | 1 | 44 |
| 463 | 公园 | 1 | 1 | gōngyuán | 1 | 1 | 40 |
| 464 | 分 | 1 | 1 | fēn | 1 | 1 | 40 |
| 465 | 甜 | 1 | 1 | tián | 1 | 1 | 27 |
| 466 | 叔叔 | 1 | 1 | shūshu | 1 | 1 | 43 |
| 467 | 青蛙 | 1 | 1 | qīngwā | 1 | 1 | 43 |
| 468 | 帮忙 | 1 | 1 | bāngmáng | 1 | 1 | 43 |
| 469 | 哦 | 1 | 1 | 0 | 1 | 43 |
III. Which words to select? Why

- Nouns, verbs, adjectives...
- Use words to express learning?
- Can dos?
  - Identify
  - Compare
  - Answer comprehension questions
  - Write
  - Retell
  - ...
How to organize?

Word cluster
How to organize?

Word cluster

蝌蚪

黑

游泳

小

卵

池塘

K-5 Chinese Word Frequency Dictionary for L2 Learners
How to organize?
Word Bank/Chart

<table>
<thead>
<tr>
<th>NAME</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>小蝌蚪</td>
<td>小、黑、卵、池塘</td>
</tr>
<tr>
<td>鸭子</td>
<td>游泳</td>
</tr>
<tr>
<td>红鲤鱼</td>
<td>大眼 鳃、嘴巴宽大、游泳</td>
</tr>
<tr>
<td>乌龟</td>
<td>四条腿、游泳</td>
</tr>
<tr>
<td>鹅</td>
<td>白肚皮、游泳</td>
</tr>
<tr>
<td>青蛙</td>
<td>四条腿、白肚皮、大眼睛、扁嘴巴 巴、 绿、游泳</td>
</tr>
</tbody>
</table>

K-5 Chinese Word Frequency Dictionary for L2 Learners
IV. After reading the story, use word bank/graphic organizer to create sentences (interpretative and presentational task)

- 蝌蚪很小。(蝌蚪________。)
- 鸭子会游泳。(_______会游泳)
- 红鲤鱼____________。
- 乌龟有四条腿。
- 鹅的肚皮是白的。
- 青蛙...绿色...
Describe frogs with appropriate words
### Attribute chart

<table>
<thead>
<tr>
<th></th>
<th>游泳</th>
<th>大眼睛</th>
<th>四条腿</th>
<th>白肚皮</th>
<th>绿色</th>
<th>黑色</th>
</tr>
</thead>
<tbody>
<tr>
<td>蝌蚪</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>鹅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>红鲤鱼</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>乌龟</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>鹅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>青蛙</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

K-5 Chinese Word Frequency Dictionary for L2 Learners
使用问题来评估理解力

- 谁会游泳？
- 谁有四条腿？谁有两条腿？谁没有腿？
- 谁的肚皮是白的？
- 谁有大眼睛、扁嘴巴？
- 谁穿黑／绿衣服？
- 谁在小时候和长大以后不一样？
春天来了，青蛙妈妈睡醒了，跳进池塘，生下很多圆圆黑黑的卵。
后来，卵变成了小蝌蚪。
他们看见小鸭子跟着鸭妈妈游水，也要去找妈妈。鸭妈妈告诉小蝌蚪：“你们的妈妈有两只大眼睛，嘴巴又宽又大。小蝌蚪高高兴兴地去找妈妈。

*Prediction: Who will be the next to ask?*
小蝌蚪看见红鲤鱼，直叫 “妈妈”，红鲤鱼却说： “你们的妈妈有四条腿。”
小蝌蚪又叫乌龟“妈妈”，乌龟说：“你们的妈妈肚皮是白的。”
小蝌蚪又遇见大白鹅，直叫“妈妈”，大白鹅说：“你们的妈妈穿着绿衣服。”
一会儿，青蛙妈妈游来了，小蝌蚪甜甜的喊“妈妈”。可小蝌蚪一看，妈妈有四条腿，觉得自己一点也不像妈妈，青蛙妈妈说：“你们还没长大。”小蝌蚪跟着妈妈做游戏戏水，慢慢地也长出了四条腿，他们长大了。
2016 Infrastructure Grant: Using the K-5 Word Frequency Dictionary to Build Vocabulary and Comprehension Assessments

Phase I    Develop character and comprehension vertical assessments aligned with the K-5 WFD and CCSS

Phase II   Field Testing, Application and Reliability Testing

Phase III  Compile and Summarize Interpretive Mode assessments developed by students’ literacy proficiency through the application of the WFD in the classroom

*Pilot Assessments and WFD: info@MandarinInstitute.org*
“It’s called ‘reading’. It’s how people install new software into their brains.”