

国际中文教育 词汇分析工具

使用说明

2025 年 6 月

中国 北京

目录

一、主要功能	1
二、 使用方法	1
三、操作步骤	1
1.打开程序	1
2. 软件操作	3
2.1 教材生词重现率统计	3
2.2 词汇-大纲对比	8
2.3 词表收词比对	11

一、主要功能

WordAnalysis 是一款词汇分析工具，适用于国际中文教育学科研究，能够统计生词在课文的重现率，统计词表在词汇大纲的位置和分布、对比词表与词表之间的词汇分布。

二、使用方法

安装程序后，根据研究目的选择上方选项卡中的功能，上传生词表和分词后的课文，可以根据某课生词表查看某篇课文中词汇的复现率；上传将要统计的词表，将词表的词汇与目前出台的大纲进行对比，查看在各个大纲词表的等级分布，从而推测词表的难度分布；导入不同词表，对比统计所研究的不同词表中的共有词和特有词。

三、操作步骤

1.打开程序

①解压安装程序压缩包→②选择“WordAnalysis-Setup”应用程序。

名称	修改日期	类型	大小
词表比对测试用：三本教材生词表	2024/8/5 23:22	文件夹	
词汇大纲比对测试用：《基础实用商务汉语》生...	2024/8/5 23:22	文件夹	
重现率统计测试用：《公司汉语》课文分词语料	2024/8/5 23:22	文件夹	
重现率统计测试用：《公司汉语》生词表	2024/8/5 23:22	文件夹	
 WordAnalysis-Setup	2025/6/11 16:47	应用程序	10,744 KB

图 1 压缩包解压界面

③按照下列流程进行程序的安装。

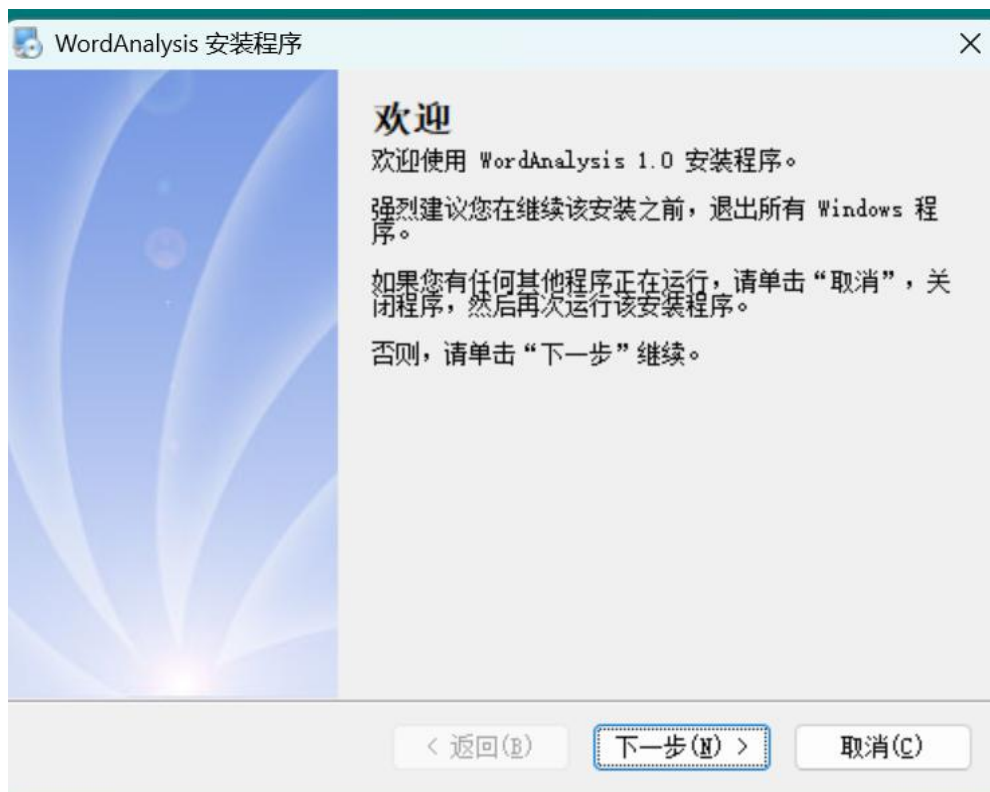


图 2 安装程序欢迎页面

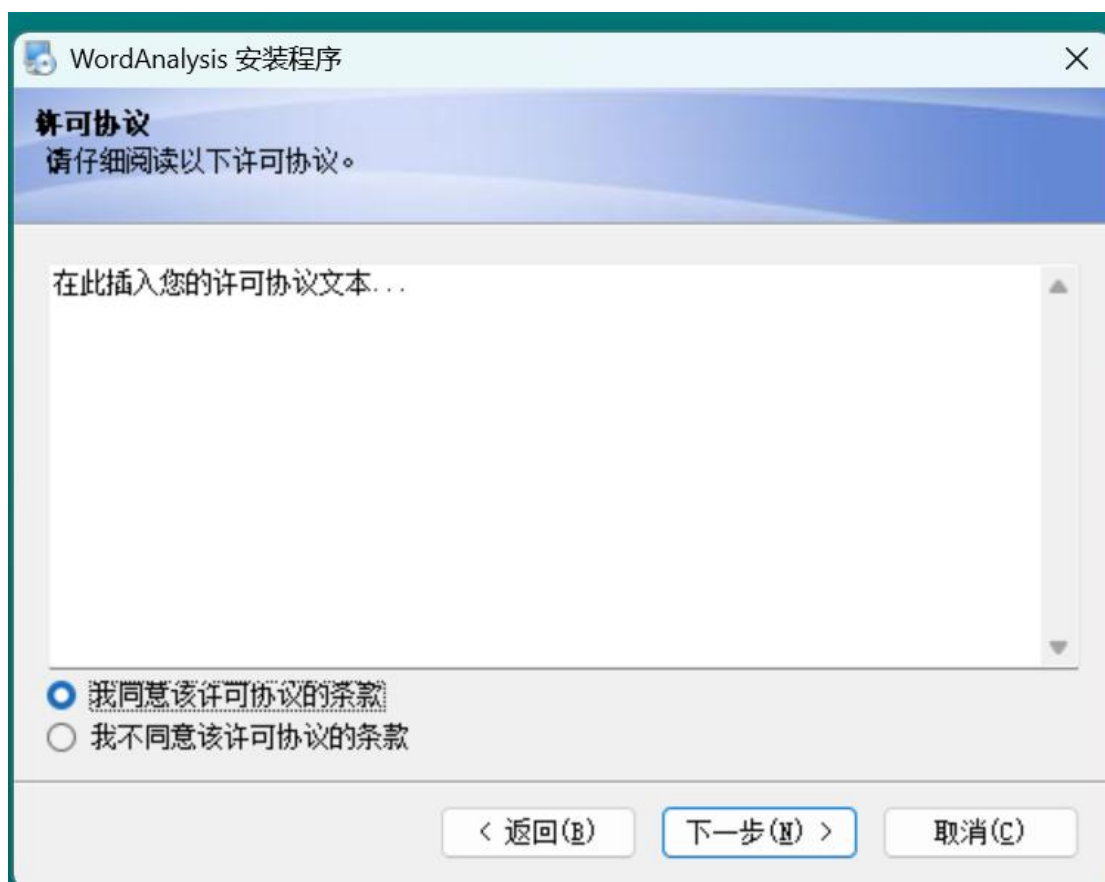


图 3 程序许可协议



图 4 程序安装位置

④在安装目录下面找到 WordAnalysis 程序并打开。

2. 软件操作

2.1 教材生词重现率统计

将生词表中的词汇在已分词的课文语料中的次数进行统计, 计算出生词在文章中的复现率, 从而有助于中文教师词汇教学以及教材编写者的生词设计。



图 5 教材生词重现统计

①处理课文，将课文语料进行分词。词与词之间以词性标注为界，若未有词性标注，则以做斜线“/”为界限。

分词后语料参考：

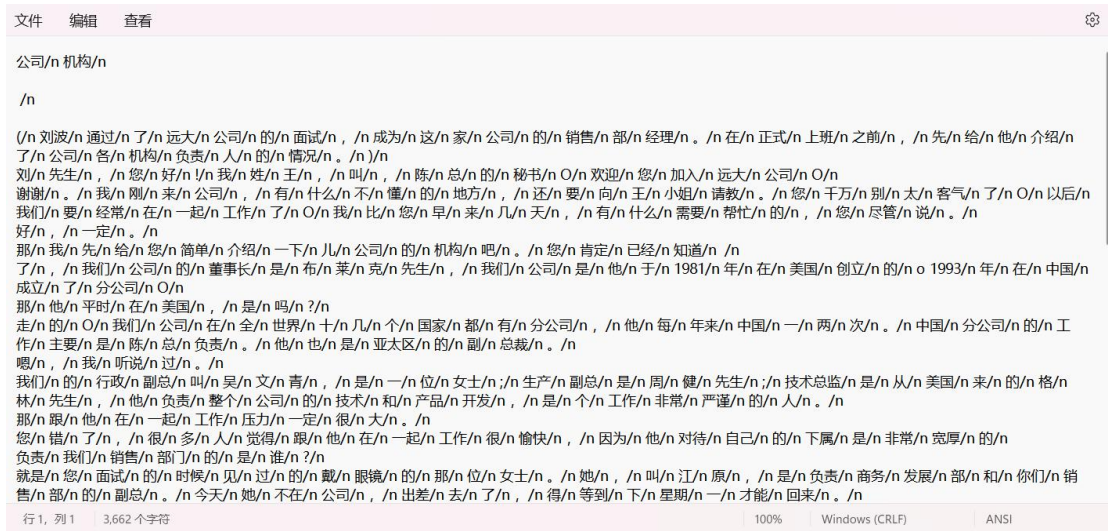


图 6 分词后的课文语料

②建立教材生词表。建立 dbf 或 txt 格式生词表，生词表中应含有词语列和课文序号列，txt 格式的生词表以“Tab”键间隔。

txt 格式的生词表参考：

文件	编辑	查看
病毒 2		
播放 7		
博览会	10	
补偿 9		
不妨 16		
不光 16		
不谋而合 5		
不相上下 4		
步伐 6		
材料 9		
财富 12		
财务 1		
采购 1		
采摘 12		
操作平台 2		
层次 6		
查毒 2		
差事 16		
长远 15		
尝试 5		
超额 15		
超越 4		
闯劲 5		
趁 12		
成 3		
行 51, 列 3	3,172 个字符	

图 7 生词表

③选择要进行重现率统计的课文数量“单篇/文件夹”，这里以单篇为例。

④选择课文语料类型“词性标注/未标注”这里以词性标注后的课文语料为例。

⑤选择已分词文件。以《公司汉语》第一课课文为例。

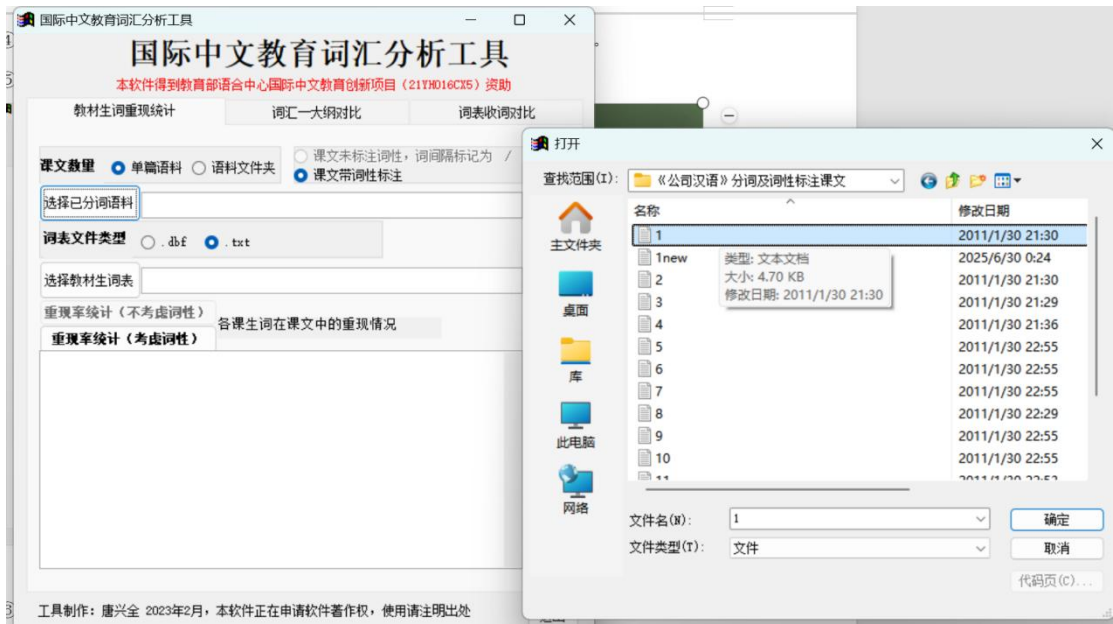


图 8 教材生词重现统计 1

⑥选择词表的文件类型

⑦选择教材生词表

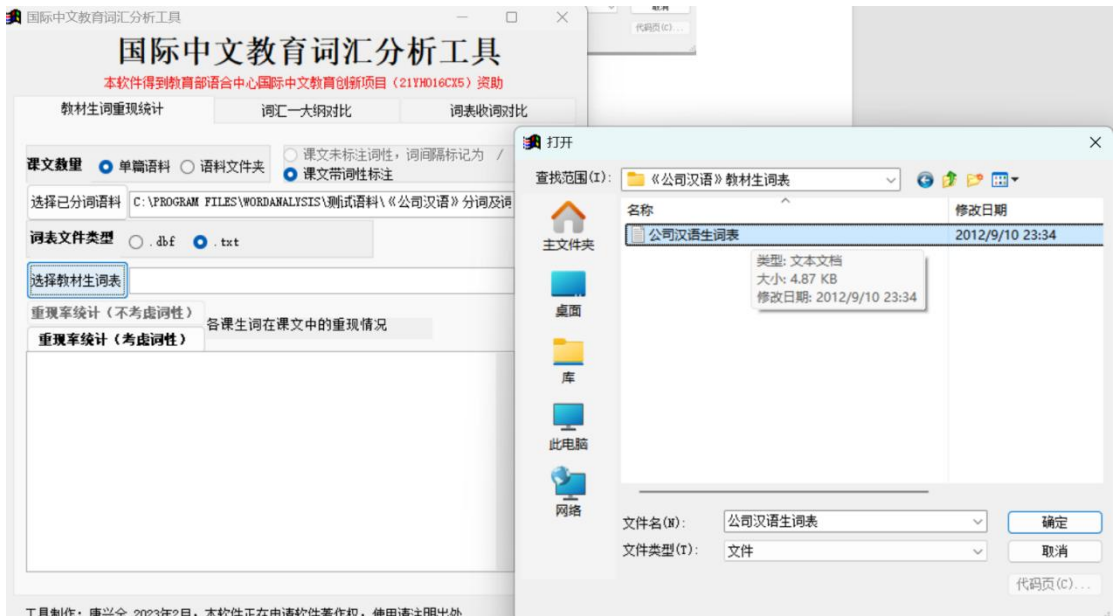


图 9 教材生词重现统计 2

⑧点击“重现率统计”

⑨观察统计结果。

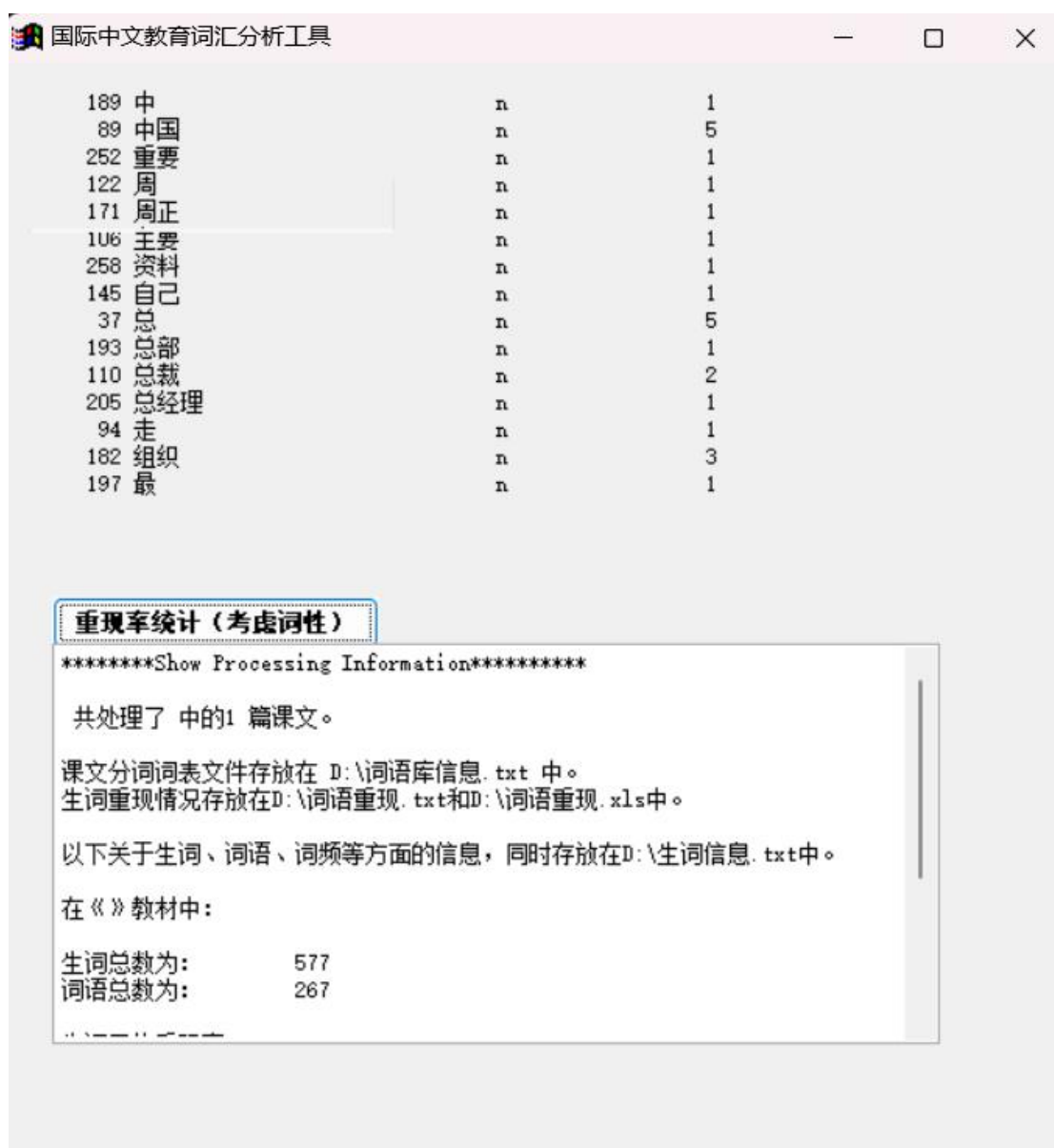


图 10 教材生词重现统计结果

在统计结果中，显示以下内容

- ①处理文章数量和处理结果保存位置。
- ②显示处理生词和词语数量
- ③生词平均重现率
- ④词语平均词频
- ⑤重现率为 0 的生词数量
- ⑥只出现了一次的词语数量

2.2 词汇-大纲对比

目前,国际中文教育领域已出台三版国际中文教育词汇等级大纲。分别是《国际中文教育中文水平等级标准词语表》、《原 HSK 词汇大纲》、《新 HSK 词汇大纲》、《BCT (商务汉语考试) 词语表》。通过将词表的词汇与大纲进行对比,可以查看在词表的等级分布,从而推测词表的难度分布。



图 11 词汇-大纲对比界面

①选择要对比的词表。以“基础实用商务汉语生词表”为例,查看该生词表在四部词汇大纲中的等级分布。

注意: ①词表内词汇排列应一行一词呈纵向排列。②采用 ANSI 编码类型。



图 12 词汇-大纲对比操作 1

②点击“比对”，选择文件保存位置

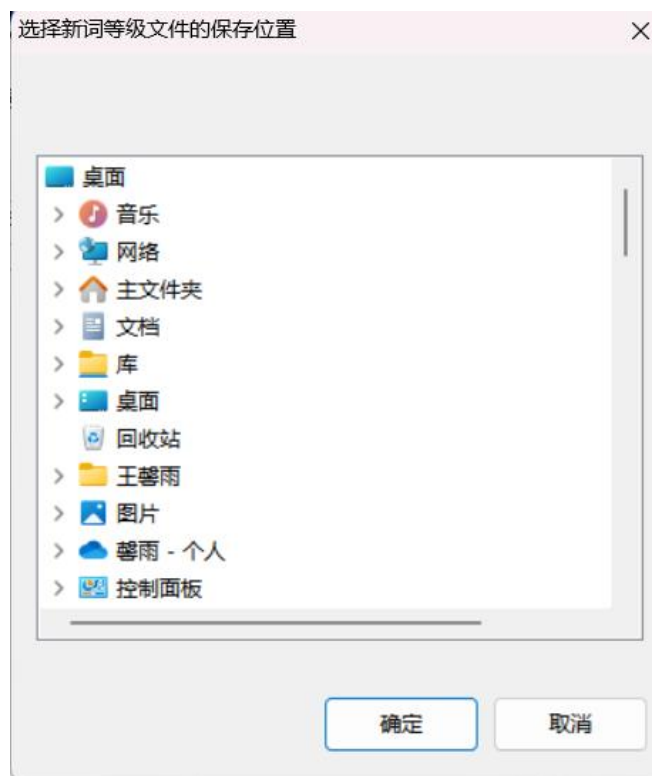


图 13 词汇-大纲对比界面 2

③查看结果。

-文件保存的位置信息。

-词表中词语总数。

-四个词表的比对结果。

-不属于四部词表大纲任一大纲的词汇。



图 14 词汇-大纲对比结果

2.3 词表收词比对

对比不同的词表, 查看收词差异。



图 15 词表收词对比界面

- ①建立词表，词表的格式应为 txt 或 dbf 格式。txt 格式的词表内部，按照一行一词纵向排列；dbf 格式应包含“ciyu”字段。
- ②选择要对比的词表数量，以 3 份词表为例。
- ③分别选择词表一、词表二、词表三的文件。



图 16 词表收词对比

④点击收词比对，选择文件保存位置。

⑤查看结果：

- 三个词表共选词数量。
- 词表一基础实用商务汉语特有词数量。
- 词表二金桥汉语特有词数量。
- 词表三经理人汉语特有词数量。
- 只出现在一个词表中的生词

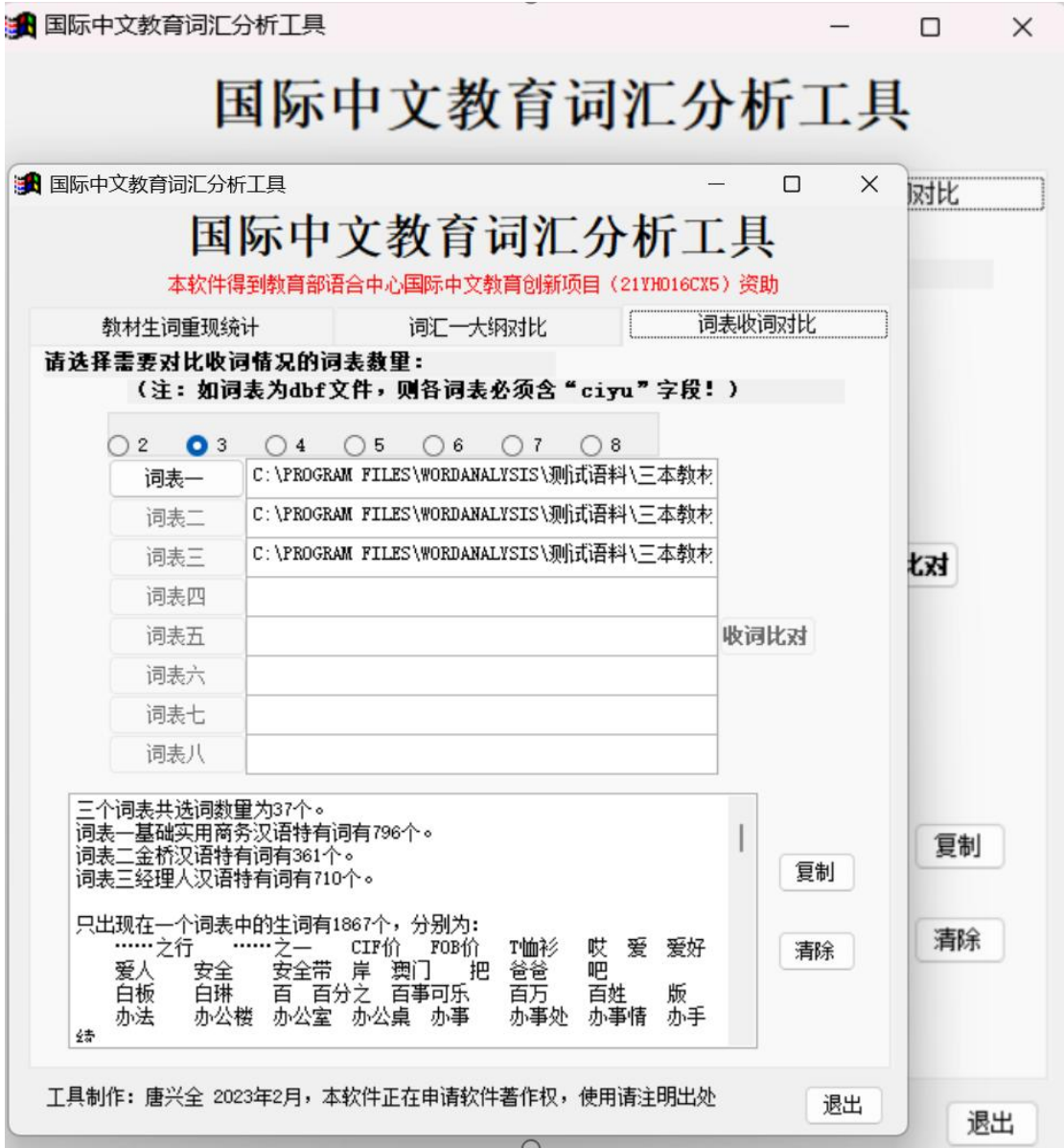


图 17 词表收词对比结果

⑥点击“退出”退出程序