

# 来华留学生 课程作文语料库 系统

## 使用说明

对外经济贸易大学

2025 年 6 月 北京

## 目 录

|                        |    |
|------------------------|----|
| 一、主要功能 .....           | 1  |
| 二、 使用方法 .....          | 1  |
| 三、操作步骤 .....           | 1  |
| 3.1 安装和登录平台 .....      | 1  |
| 3.2 平台操作 .....         | 3  |
| 3.2.1 语料的检查和文件格式 ..... | 3  |
| 3.2.3 语料文库和句库的生成 ..... | 11 |
| 3.2.4 语料信息的检索 .....    | 16 |
| 3.2.5 语料的加工 .....      | 22 |
| 3.2.6 语料显示 .....       | 34 |
| 3.2.7 帮助 .....         | 34 |

# 一、主要功能

WritingCorp 是来华留学生课程作文语料库操作平台。该平台一站式集合语料的生成、检索、加工工具，实现来华留学生课程作文语料库在字、词、句层面的处理和分析等功能。

# 二、使用方法

安装软件后，选择登陆身份，即可进入该软件，通过点击上方不同的选项卡来实现不同的功能需求。

# 三、操作步骤

## 3.1 安装和登录平台

①解压 WritingCorp 压缩包,找到并点击“WritingsCorp”即可打开应用程序，进入登录界面。

| 名称              | 修改日期            | 类型         | 大小       |
|-----------------|-----------------|------------|----------|
| DATA BAT        | 2025/6/30 0:14  | 文件夹        |          |
| 测试语料            | 2025/6/28 17:24 | 文件夹        |          |
| CIBIAO          | 2017/3/30 23:55 | CDX 文件     | 615 KB   |
| cibiao.dbf      | 2017/3/30 23:55 | DBF 文件     | 688 KB   |
| comdlg32.ocx    | 2004/3/9 16:45  | ActiveX 控件 | 150 KB   |
| config.fpw      | 2006/3/15 8:49  | FPW 文件     | 1 KB     |
| FOXUSER.DBF     | 2025/6/30 0:14  | DBF 文件     | 1 KB     |
| FOXUSER.FPT     | 2025/6/30 0:14  | FPT 文件     | 1 KB     |
| mscomct2.ocx    | 2004/3/9 16:45  | ActiveX 控件 | 647 KB   |
| MSVCP71.DLL     | 2003/3/19 11:14 | 应用程序扩展     | 488 KB   |
| msvcr71.dll     | 2004/1/12 8:00  | 应用程序扩展     | 340 KB   |
| VFP9CHS.DLL     | 2004/12/14 7:57 | 应用程序扩展     | 1,704 KB |
| VFP9CHT.DLL     | 2004/12/14 7:57 | 应用程序扩展     | 1,704 KB |
| VFP9ENU.DLL     | 2004/12/14 3:21 | 应用程序扩展     | 1,704 KB |
| VFP9KOR.DLL     | 2004/12/14 7:57 | 应用程序扩展     | 1,704 KB |
| vfp9r.dll       | 2005/11/5 10:13 | 应用程序扩展     | 4,612 KB |
| vfp9rchs.dll    | 2005/11/5 10:15 | 应用程序扩展     | 1,104 KB |
| VFP9RENU.DLL    | 2005/11/5 9:44  | 应用程序扩展     | 1,160 KB |
| vfp9t.dll       | 2005/11/5 10:13 | 应用程序扩展     | 3,800 KB |
| writingcorp.app | 2025/6/5 0:38   | APP 文件     | 2,665 KB |
| WritingsCorp    | 2025/6/5 0:38   | 应用程序       | 2,674 KB |

图 1 安装界面一

②账号切换至“普通用户”点击确定，即登陆成功

登录

— □ ×

欢迎使用语料库平台！  
商务汉语话语标记语料库 & 来华留学生课程作文语料库

账 号

普通用户

密 码

确定

取消

图 2 登陆界面



图 3 系统界面

### 3.2 平台操作

### 3.2.1 语料的检查和文件格式

在语料选项卡之下，可实现对作文语料的检查、读入、还原和保护。



图 4 语料选项卡

### 3.2.1.1 语料检查

该功能可对作文语料中的格式和内容进行检查。

①点击“语料检查”→②选择单个文件检查或按文件夹进行批量检查

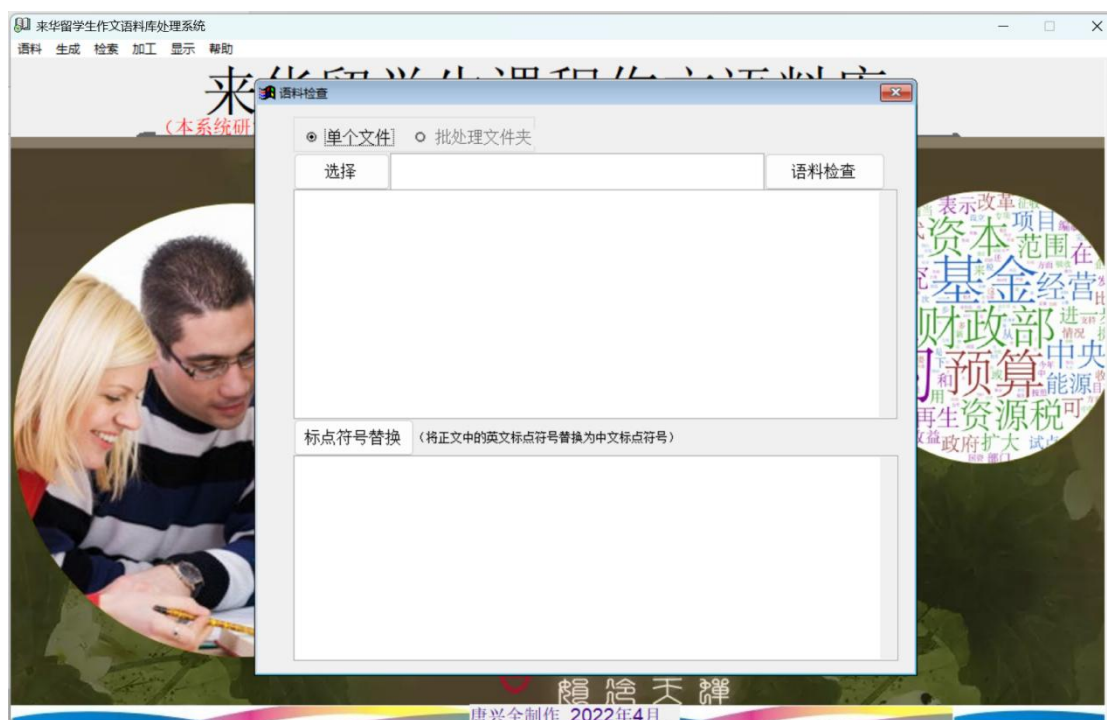


图 5 语料检查界面

③选择需要进行检查的话语标记语料，以“赢在中国.txt”语料为例。

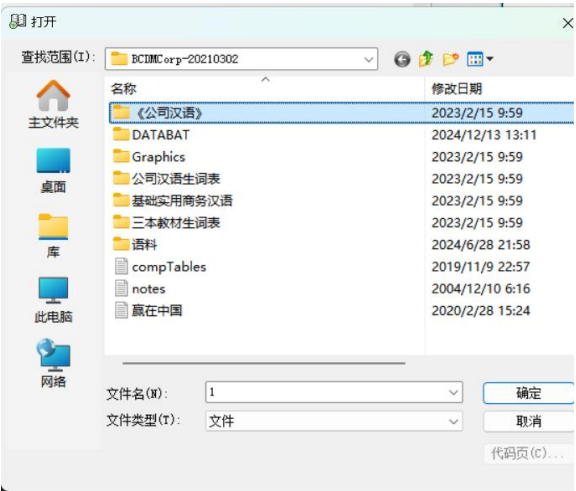


图 6 选择检查语料

④点击语料检查，观察检查结果

检查结果将从格式标注和内容标注两方面显示，从结果中可以看出，“赢在中国.txt”话语标记文件存在内容标注错误，均为标点符号类问题。可点击右侧“复制”对文件错误进行粘贴保存。

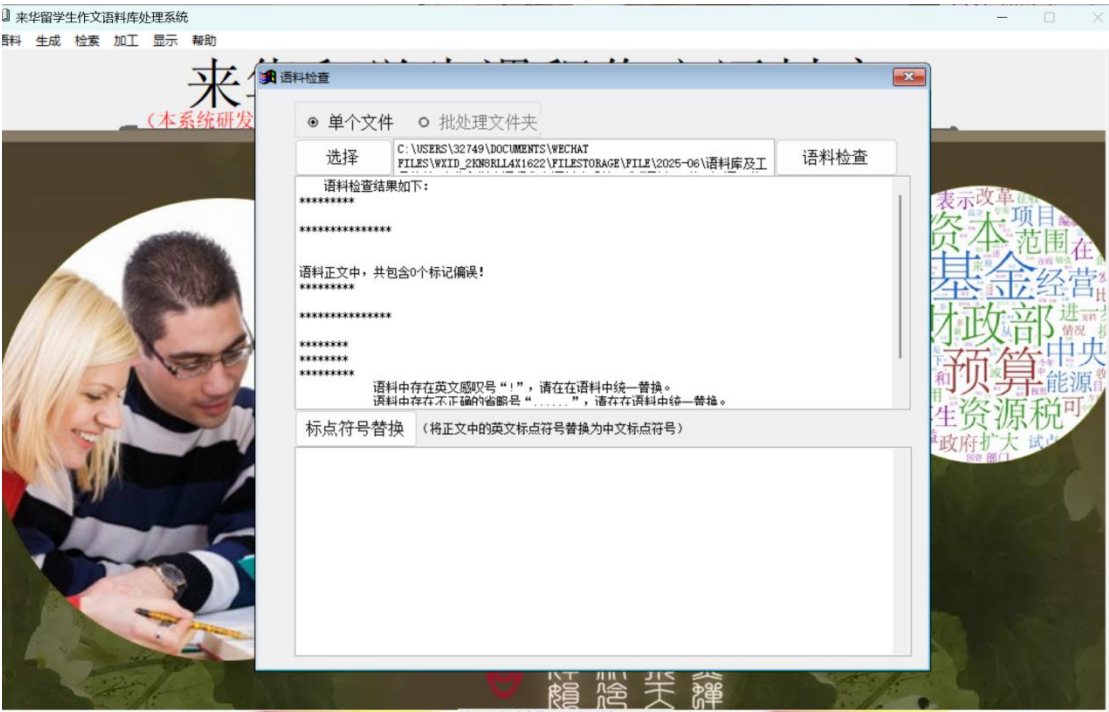


图 7 语料检查结果

解决这类问题，可以使用下方“标点符号替换”功能。

⑤ 点击“标点符号替换”，将正文中的英文标点符号转换为中文标点符号，同时将偏

误标记内容中的标点符号进行了话语标记替换。替换后新生成语料文件为\赢在中国 new.txt

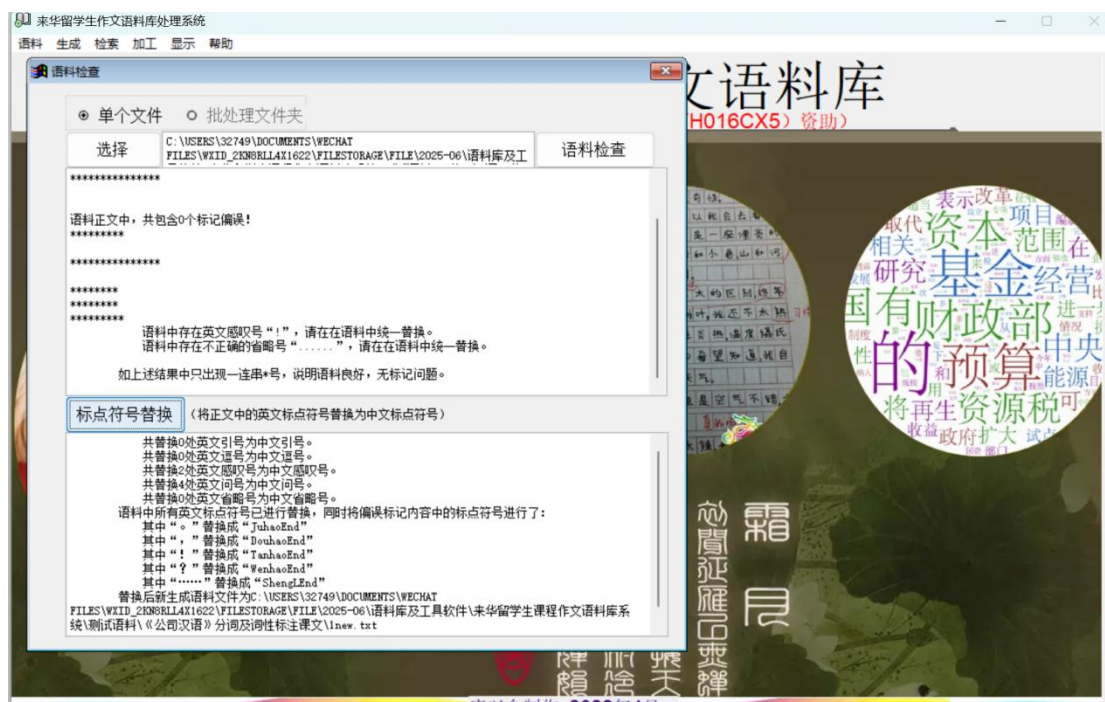


图 8 语料检查结果修正

⑥当多个话语标记类文本存放于同一文件夹之下，使用“批处理文件夹”便于提高检查效率。

### 3.2.1.2 语料读入

该功能为话语标记类语料的读入，可将.txt 格式文件读入为 XML 和 DBF 两类格式的话语标记语料文件。



图 9 语料读入选项卡

以“读入为 XML 格式”为例，读入为 DBF 格式的步骤与之相同。

①点击读入为 XML 格式。→②选择需要输出为无格式标注和内容标注问题的 txt 文件，以上述检查修改后的“赢在中国\_new.txt”为例。

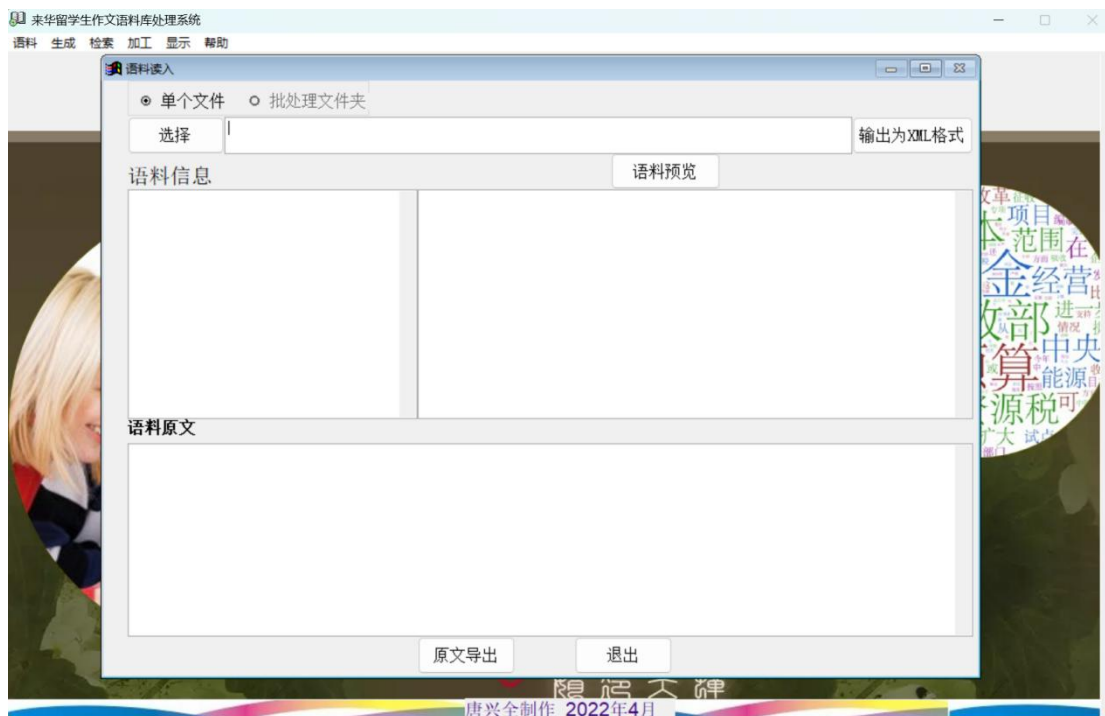


图 10 语料读入操作界面

③点击“输出为 XML 格式”，弹出 xml 格式写入成功。左上区域显示文件格式转化过

程的处理信息。

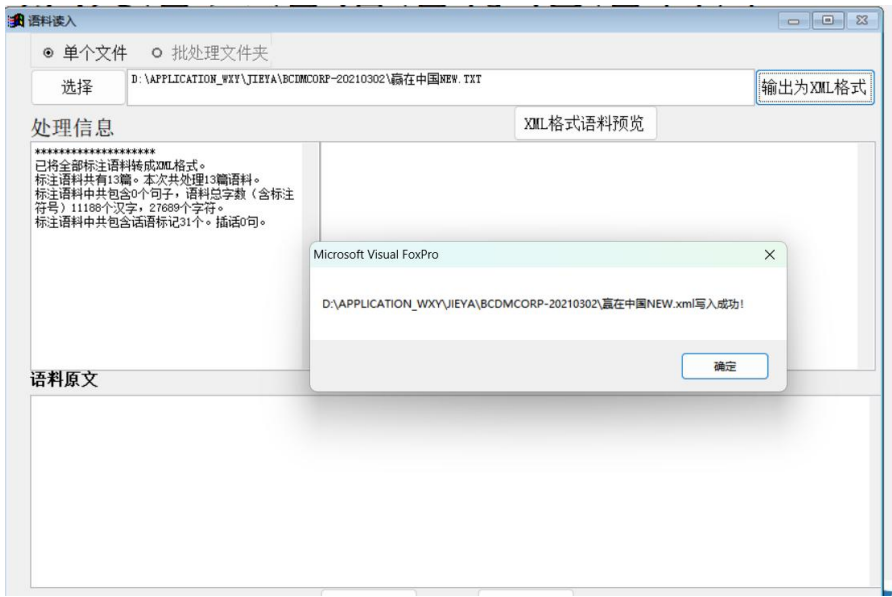


图 11 导入语料

④在该窗口下，点击“XML 格式语料预览”则可浏览 XML 格式的语料内容。



图 12 语料读入

⑤点击“批处理文件夹”，txt 格式文件批量为 XML 格式。

⑥点击“退出”，退出当前选项，并清除语料浏览痕迹。

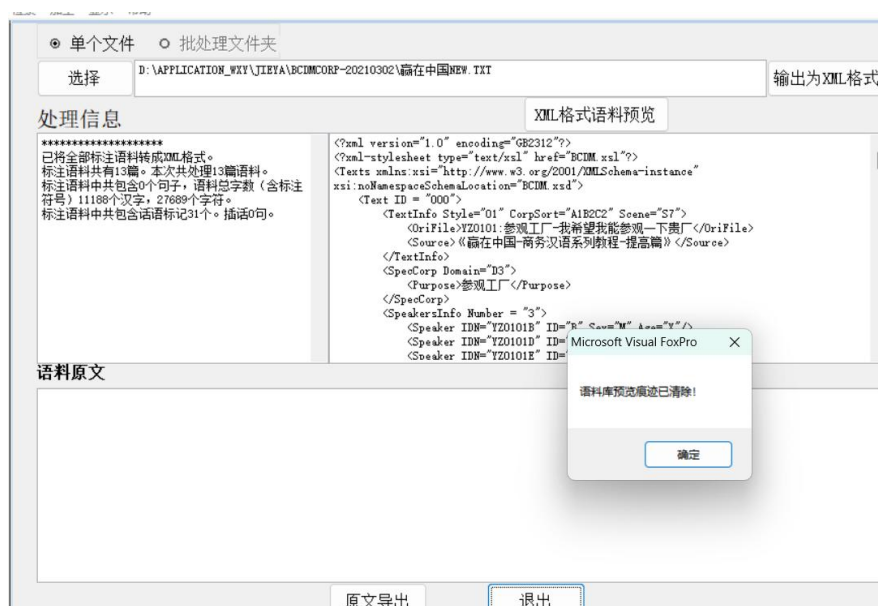


图 13 退出语料读入功能

### 3.2.1.3 语料还原

语料读入将文件格式由文本文件读入为数据库文件，此处语料还原功能，则将数据库文件还原为文本文件。



图 14 语料还原选项卡

①点击“选择数据库”，选择数据库格式的语料文件。→②点击“还原为文本”，则可

转换为 txt 格式的语料文件。

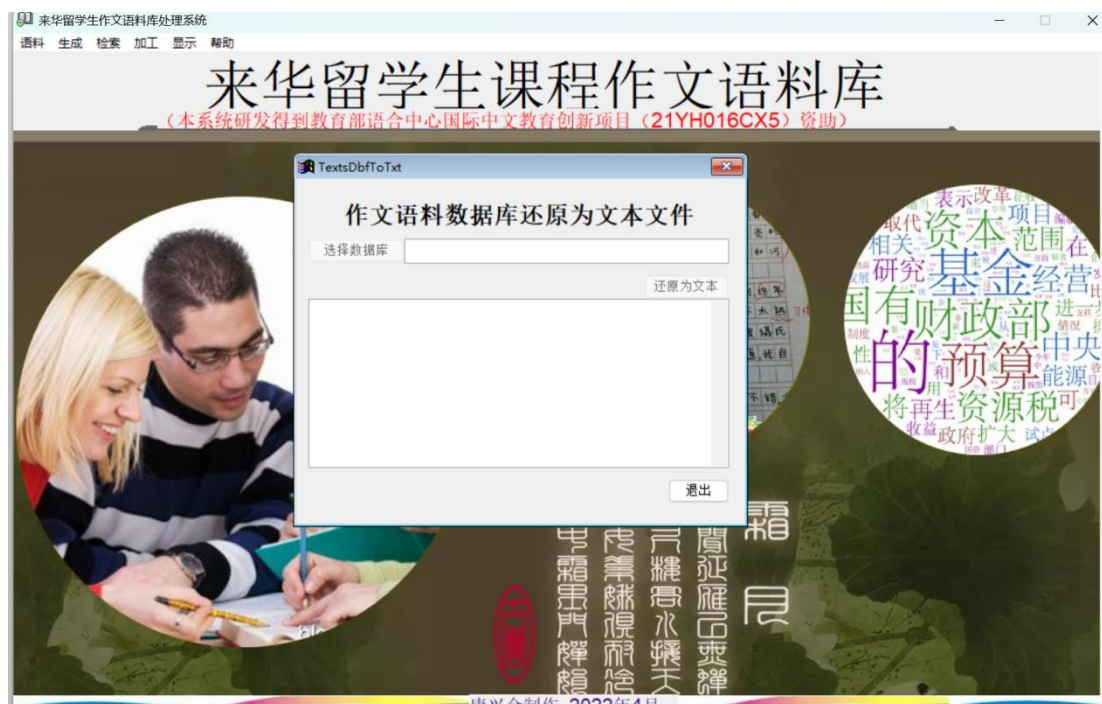


图 15 语料还原选项卡

### 3.2.1.4 语料保护

为防止真实语料出现删除、改写、增添等情况，可采取语料保护，对语料进行加密。



图 16 语料保护选项卡

①选择需要加密的文件。→②点击“加密”

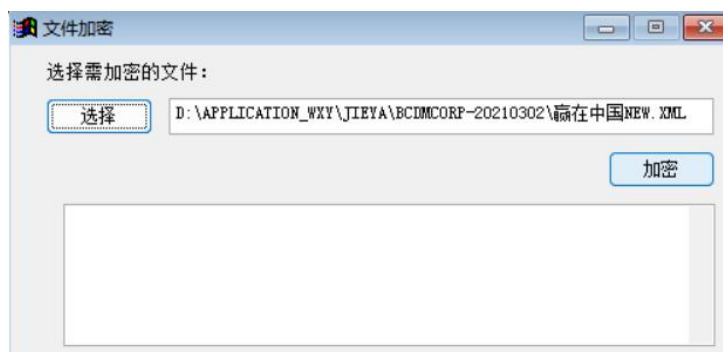


图 17 文件加密

③输入密码，加密成功。

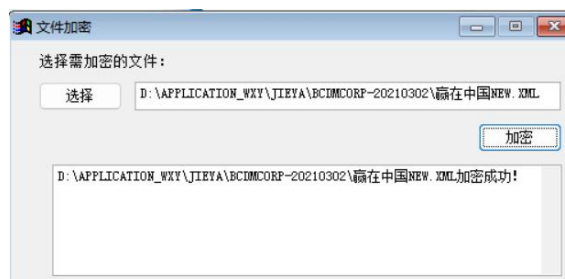


图 18 选择加密文件

### 3.2.3 语料文库和句库的生成

在“生成”的选项卡下，可以实现文库和句库的生成以及偏误统计结果的生成。



图 19 生成选项卡

### 3.2.3.1 文库生成

该功能为在总库的基础上，依据研究目的选择的语料属性和学习者属性生成子库。

①选择结合语料属性生成字库，可根据命题类型“命题/自拟/无题”、作文体裁、有无辅助、课型“商务汉语写作/经贸汉语精读/高级商务汉语/经贸汉语阅读与写作”、写作时长、作文类型等具体属性生成子库。

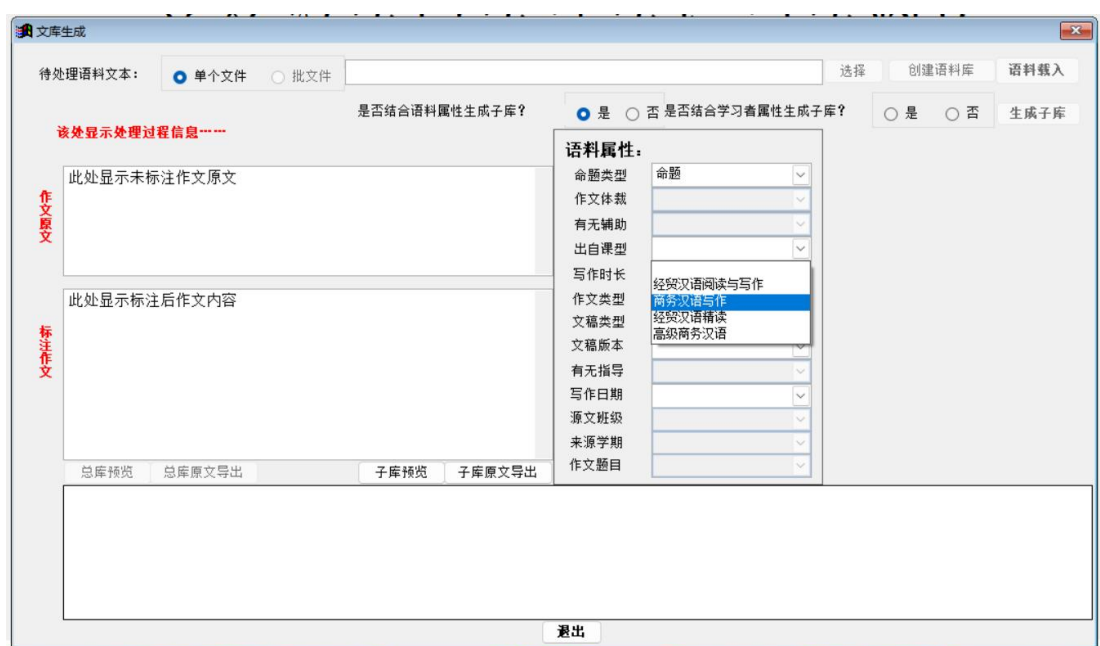


图 20 文库生成功能页面

②选择结合学习者属性生成子库，可生成根据来自不同国家、不同母语、不同汉语水平生成子库。→③点击“生成子库”，即可在总库基础上生成符合上述属性的语料库。

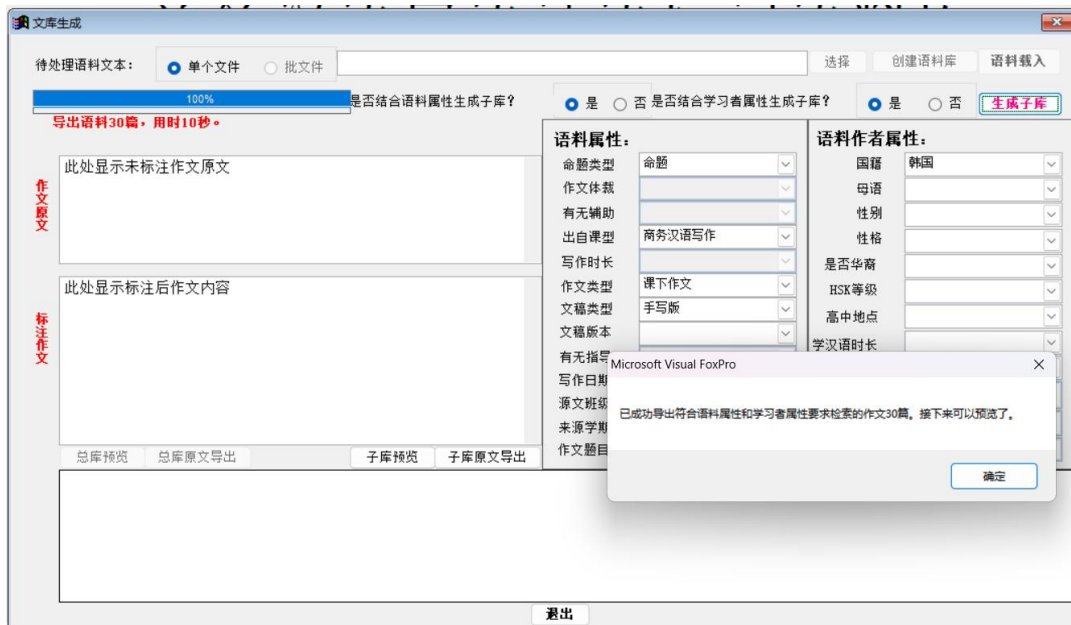


图 21 文库生成操作页面 1

④点击“子库预览”，预览满足上述属性的语料信息。

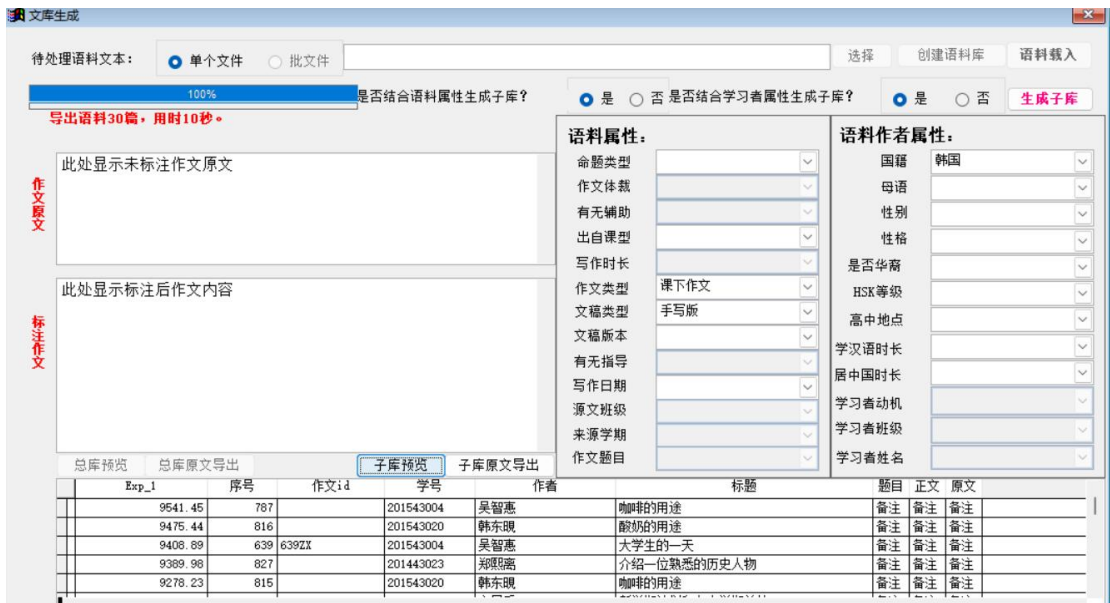


图 22 文库生成操作页面 2

⑤点击具体条目单元格，即可在左侧区域预览具体语料。

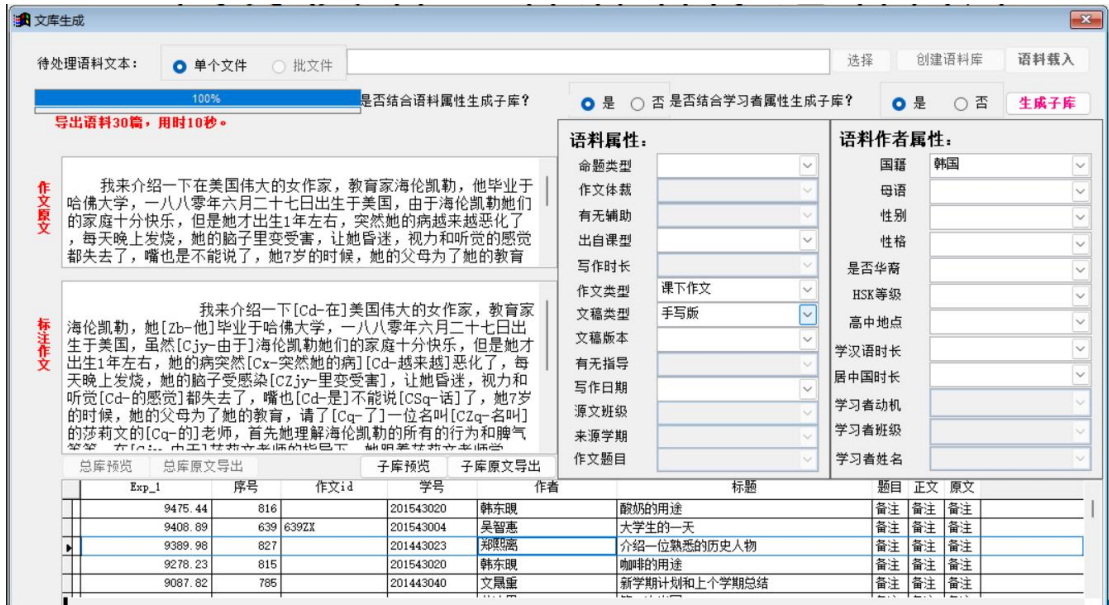


图 23 文库生成操作页面 3

⑥ “批文件” 可选择多个子语料文件，创建总库。

### 3.2.3.2 句库生成

生成以句子为单位的语料库。

① 点击“创建作文句库”，即创建句库的容器。

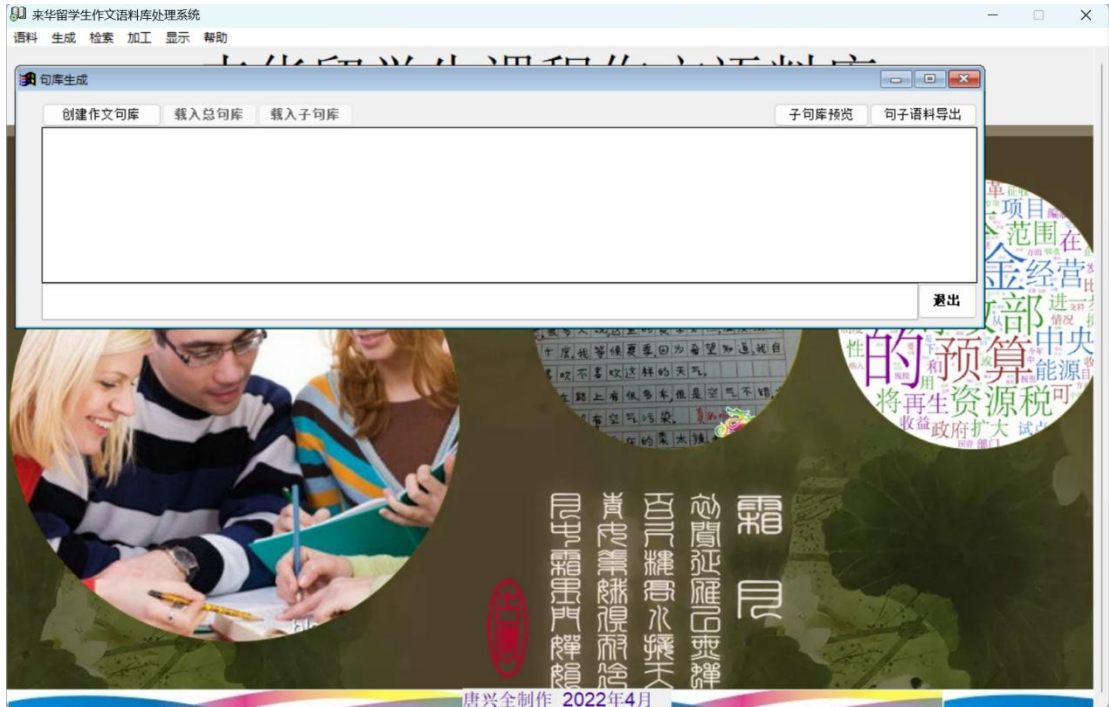


图 24 句库生成操作页面 1

② 点击“载入子句库”将文库生成操作中建立的子库以句子为单位载入。



图 25 句库生成操作页面 2

③点击“子句库预览”。点击单元格，可预览相应的句子。



图 26 句库生成操作页面 3

④点击“句子语料导出”。将预览的句子进行导出另存为保存。

⑤点击“退出”，退出该功能。

### 3.2.3.3 偏误信息统计

该功能可以对总语料库和“文库生成”过程中建立的子语料库进行偏误信息统计。本功能的实现依据对作文语料的话语标记信息的统计。点击左侧偏误信息统计类型，可查看对应统计结果。

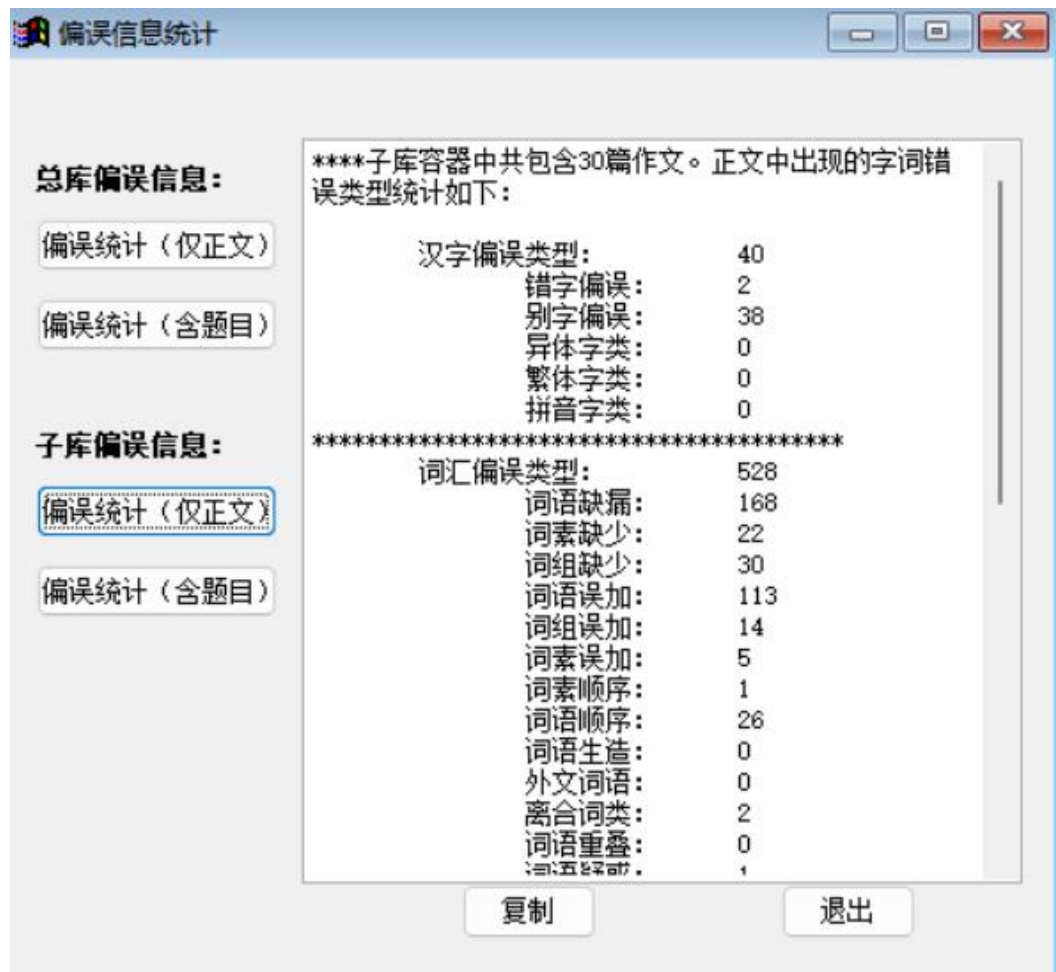


图 27 偏误信息统计

### 3.2.4 语料信息的检索

在“检索”选项卡下，可以检索偏误项、作者信息。



图 28 检索选项卡

### 3.2.4.1 话语标记的检索

对话语标记进行检索，查看含有该项标记的话语信息。



图 29 话语标记检索操作界面 1

①选择综合各方面信息检索。



图 30 话语标记检索操作界面 2

②可选择语篇特征或谈话人特征或话语标记特征或者综合三类特征检索。→③选择保存位置进行保存。

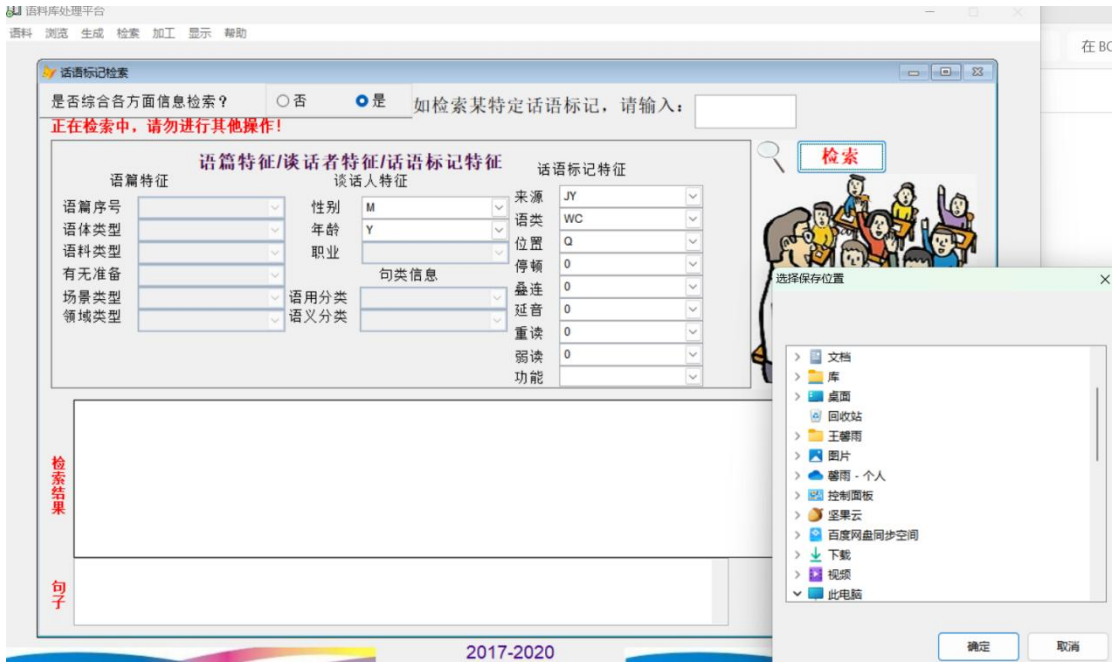


图 31 话语标记检索操作界面 3

④点击结果中的单元格，可查看具体的话语信息。



图 32 话语标记检索操作界面 4

⑥如检索某特定的话语标记，可以填入特定的话语标记信息，进行检索。

3.2.4.2 偏误项目检索

对商务汉语学习者的作文语料中标记的偏误进行分类检索。



图 33 偏误项目检索操作界面 1

①可根据研究目的选择是否结合学习者信息进行检索，以选择“韩国”学习者信息为例，选择汉字偏误类型，点击“汉字偏误检索”，进行汉字偏误中的别字检索。

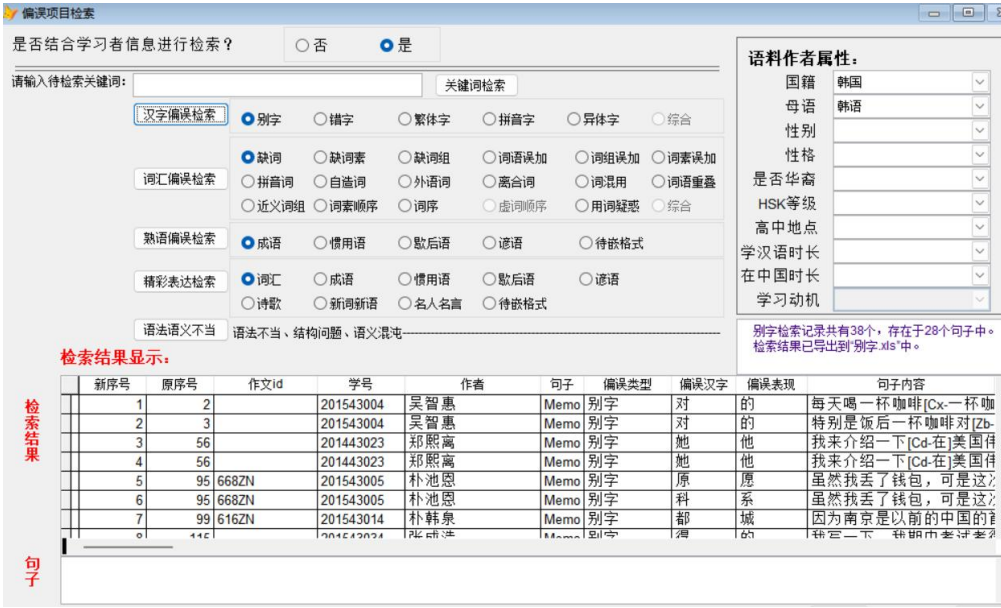


图 34 偏误项目检索操作界面 2

②选择词汇偏误类型，点击“词汇偏误检索”可进行词汇偏误中的“词序”偏误检索。  
→③选择保存位置，

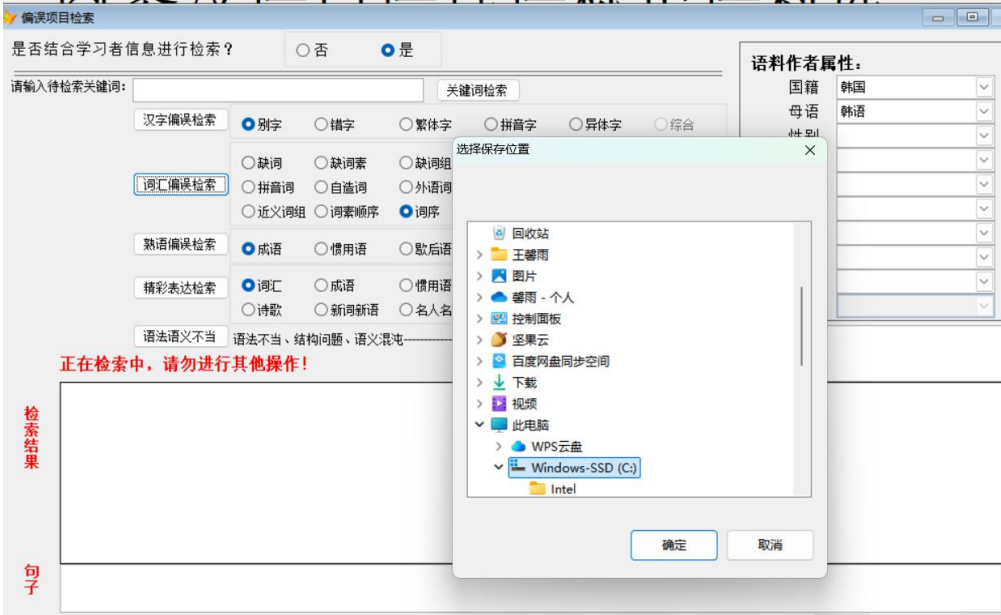


图 35 偏误项目检索操作界面 3



图 36 偏误项目检索操作界面 4

- ③熟语偏误检索、精彩表达检索、语法语义不当检索方式与上述偏误类型检索相同。
- ④支持输入语料关键词进行检索。在“请输入待检索关键词”后填入要检索的内容，点击“关键词检索”。



图 37 偏误项目检索操作界面 5

- ⑤上述检索见过均以 XLS 的格式导出至所选文件位置下。

### 3.2.4.3 作者信息检索

该功能可对作文语料的作者进行检索，便于研究者对相同国家、母语、汉语水平等作者

信息的写作材料展开历时深入研究。



图 38 作者信息检索选项卡

### 3.2.5 语料的加工

该选项卡的主要功能为对语料的字、词信息进行加工分析。



图 39 加工选项卡

### 3.2.5.1 字频统计

对语料的汉字出现频率进行统计。



图 40 字频统计操作界面 1

①选择待统计的语料文件，以公司汉语第一课为例。

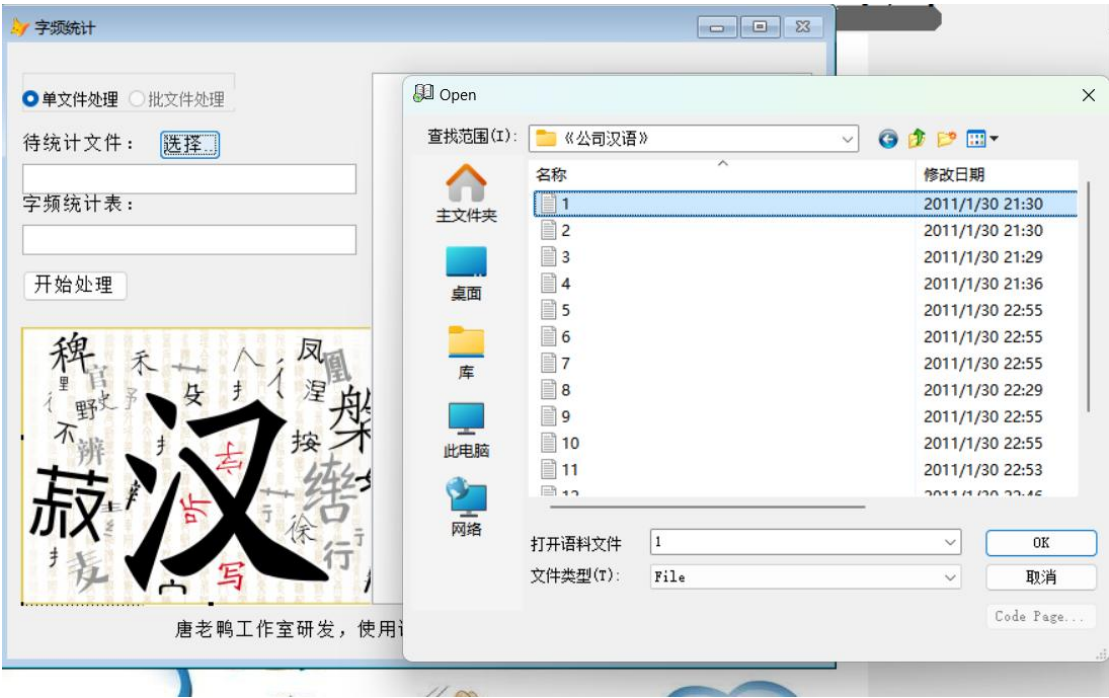


图 41 字频统计操作界面 2

②选择完毕后，设置字频统计表的保存位置和保存文件名。

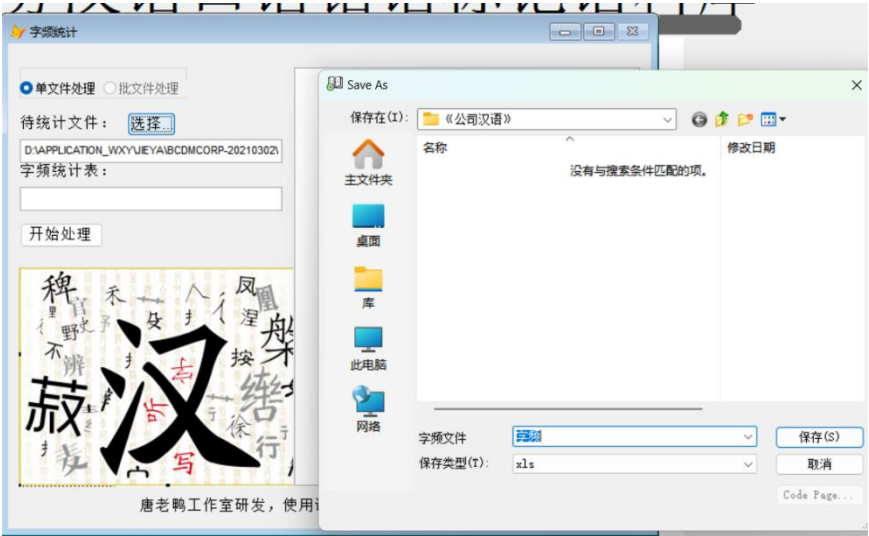


图 42 字频统计操作界面 3

③点击“开始处理”，右侧窗口显示该语料的字频统计信息。字频统计结果保存在.\《公司汉语》\字频.XLS 中。

④统计结果分为以下信息。

\*\*\*语料中共有 1028 个字。

\*\*\*字种数共有 302 个。

\*\*\*频率最高的前 10 个汉字为

\*\*\*频率最高的前 10 个汉字累积频率为

- \*\*\*频率最高的前 100 个汉字为
- \*\*\*频率最高的前 100 个汉字累积频率为
- \*\*\*频率最高的前 500 个汉字累积频率为
- \*\*\*频率最高的前 1000 个汉字累积频率为

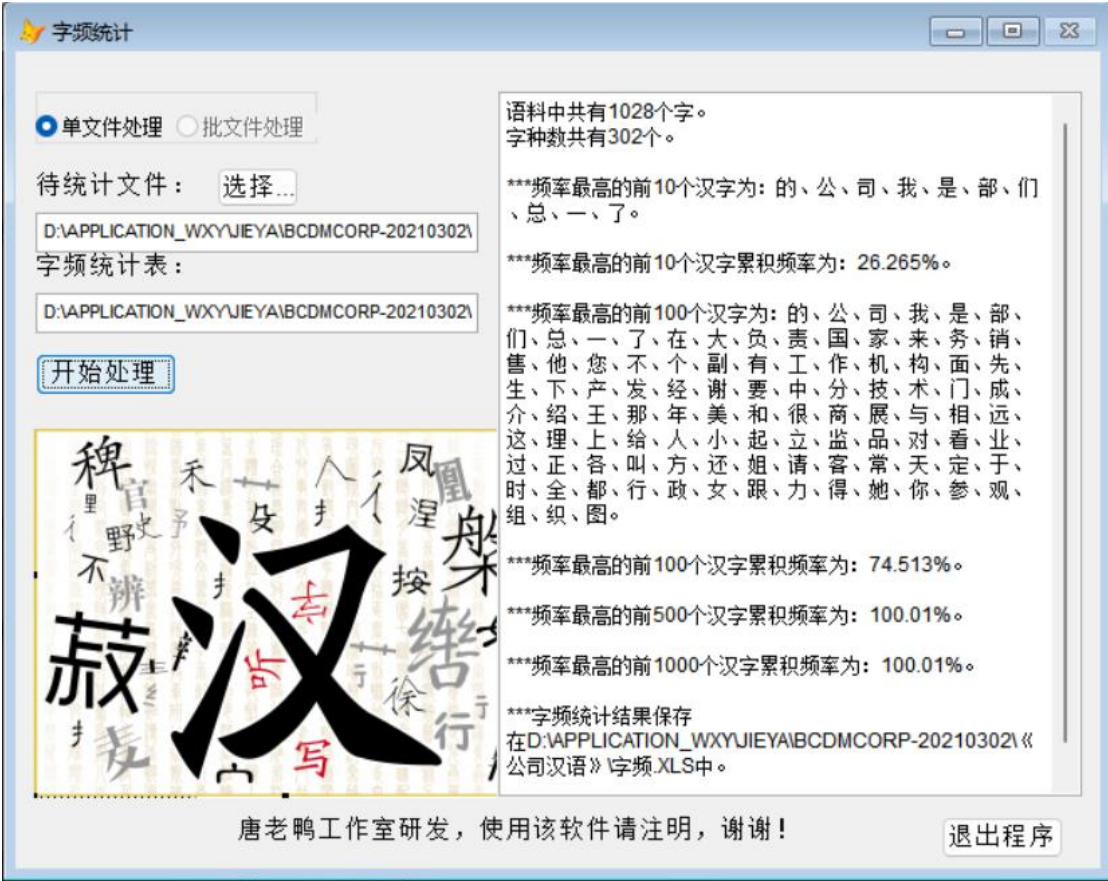


图 43 字频统计操作界面 4

3.2.5.2 分词统计

对商务汉语语料进行分词。

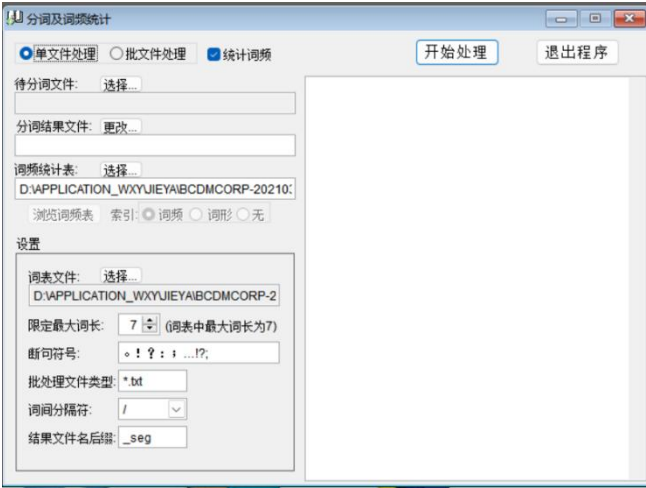


图 44 分词统计操作界面 1

①选择“单文件处理”→②选择待分词文件，以《公司汉语》第一课为例，“分词结果文件”、“词频统计表”自动填入。

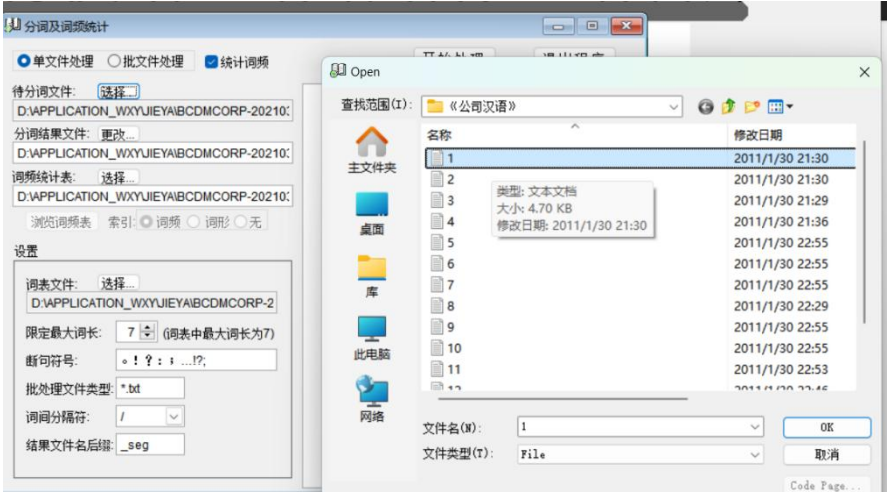


图 45 分词统计操作界面 2

②设置分词要求，词语的最大字长、句间断句符号、分词间隔符形式、以及新文件的命名后缀等。

③点击“开始处理”，选择默认文件保存路径。

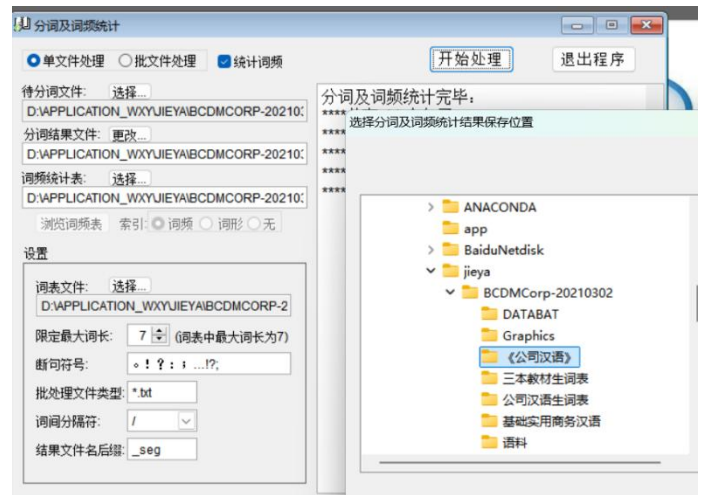


图 46 分词统计操作界面 3

④结果分析展示在右侧区域。结果主要包含以下几类信息：

- \*\*\*\*共有 XX 个句子
- \*\*\*\*共有 XXX 个字词
- \*\*\*\*总计 XXX 词次
- \*\*\*\*总计 XXX 字次
- \*\*\*\*平均句长为 XXX 个汉字

⑤词频统计结果保存在“词频表.xls”中

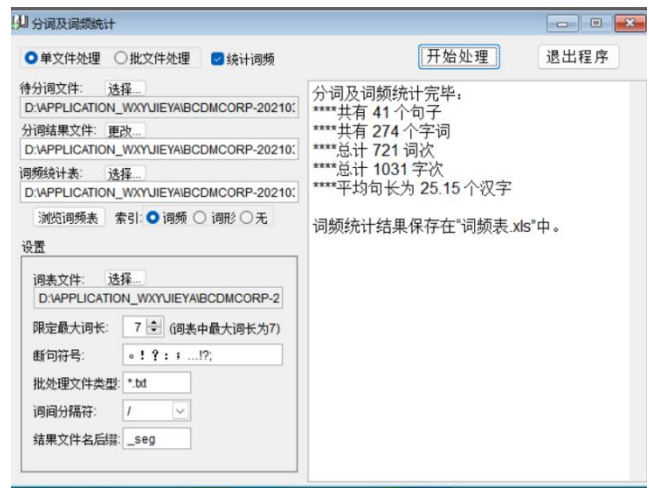


图 47 分词统计操作界面 4

⑥选择“批文件处理”实现对文件夹内所有语料文件的一键分词。以《公司汉语》文件夹为例，点击“开始处理”实现对 15 个文件的语料分词。



图 48 分词统计操作界面 5

⑦点击“退出程序”

### 3.2.5.3 词汇分析

词汇分析工具可以实现对词语重现次数的统计、词汇在大纲中的分布、不同词表的对比以及文本难度的分析。



图 49 词汇分析工具界面

#### 3.2.5.3.1 词语重现统计

①选择经过上述“分词统计”的加工语料文件。→②选择词表的文件类型。③点击“统

计重现率”可以得出词表中的词语在语料中的重现比率。

### 3.2.5.3.2 词汇-大纲对比

该功能将所选词汇与三部大纲的词汇等级进行比较。

①选择词表，以公司汉语生词表为例。

注意：文本类型的词表文件应一词一行纵向排布。

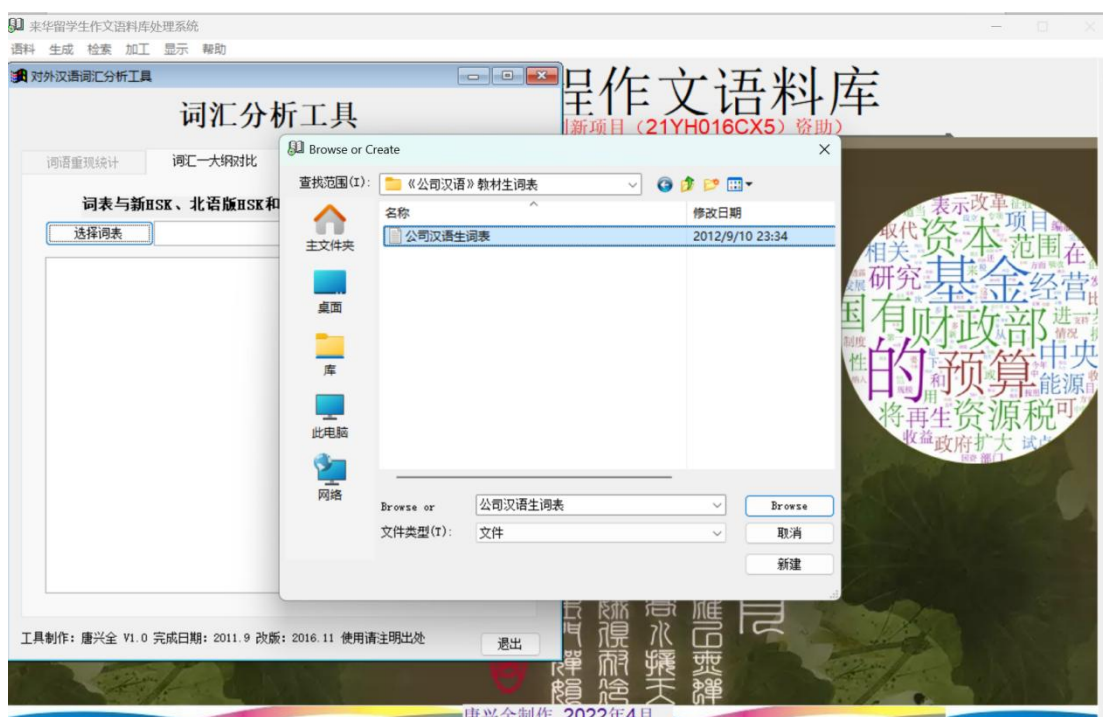


图 50 词汇分析工具操作界面 1

②点击“比对”，选择文件保存位置

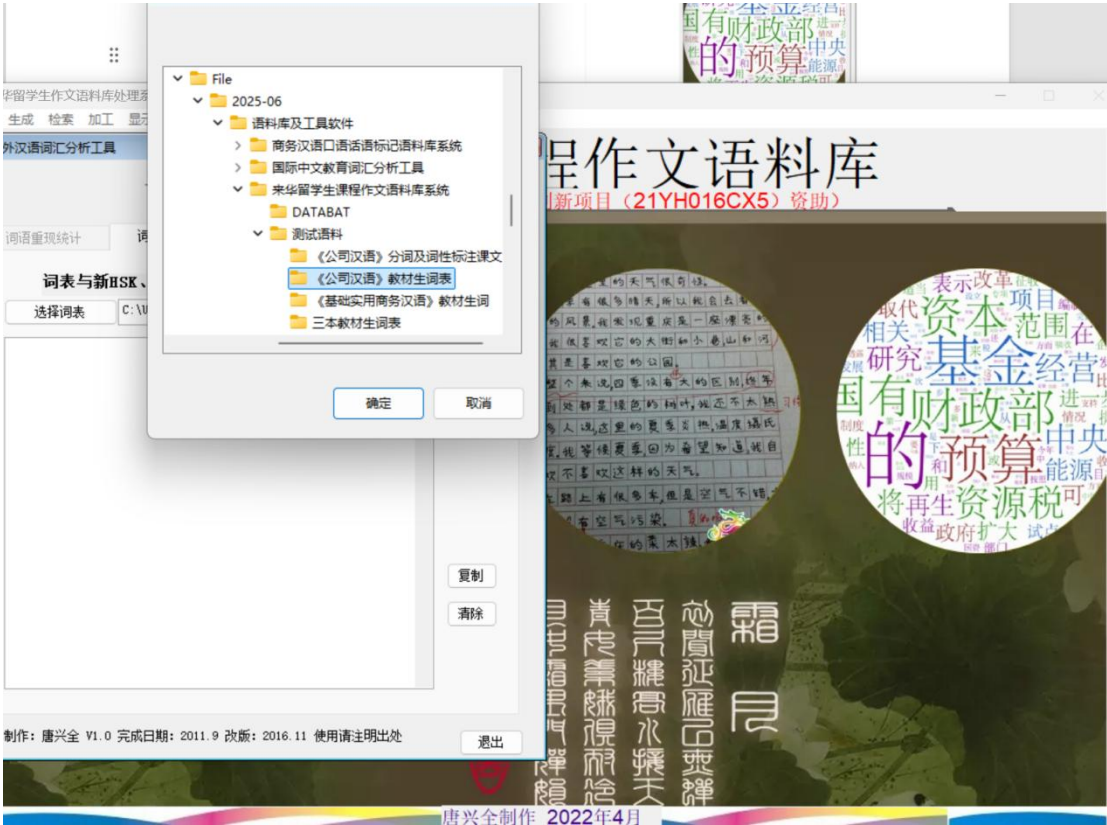


图 51 词汇分析工具操作界面 2

③在下部窗口区域显示比对结果。

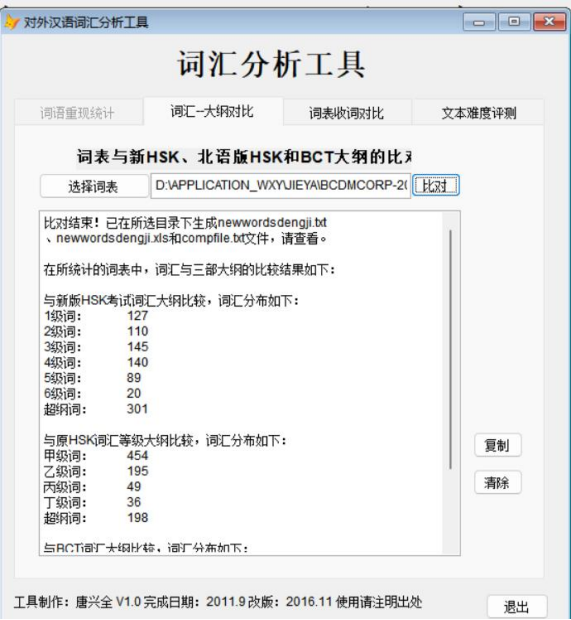


图 52 词汇分析工具操作界面 3

### 3.2.5.3.3 词表收词对比

①选择需要比对收词情况的词表数量

注意：

如为 txt 格式的文件，则“一词一行”纵向排布。如为 dbf 文件，则各词表必须含有“ciyu”字段。

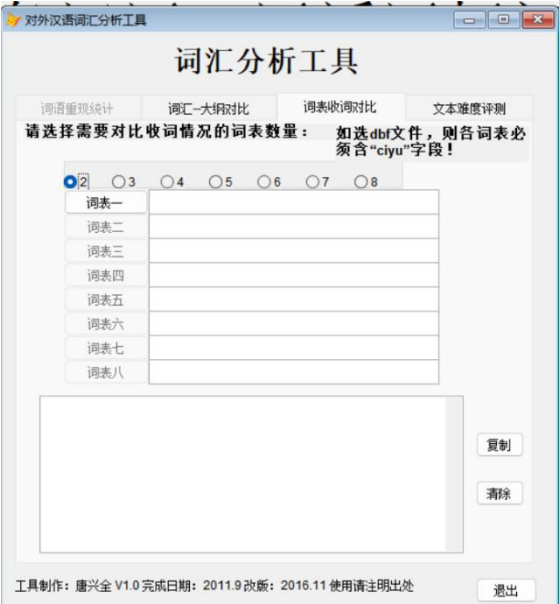


图 53 词汇分析工具操作界面 4

②选择词表的位置。

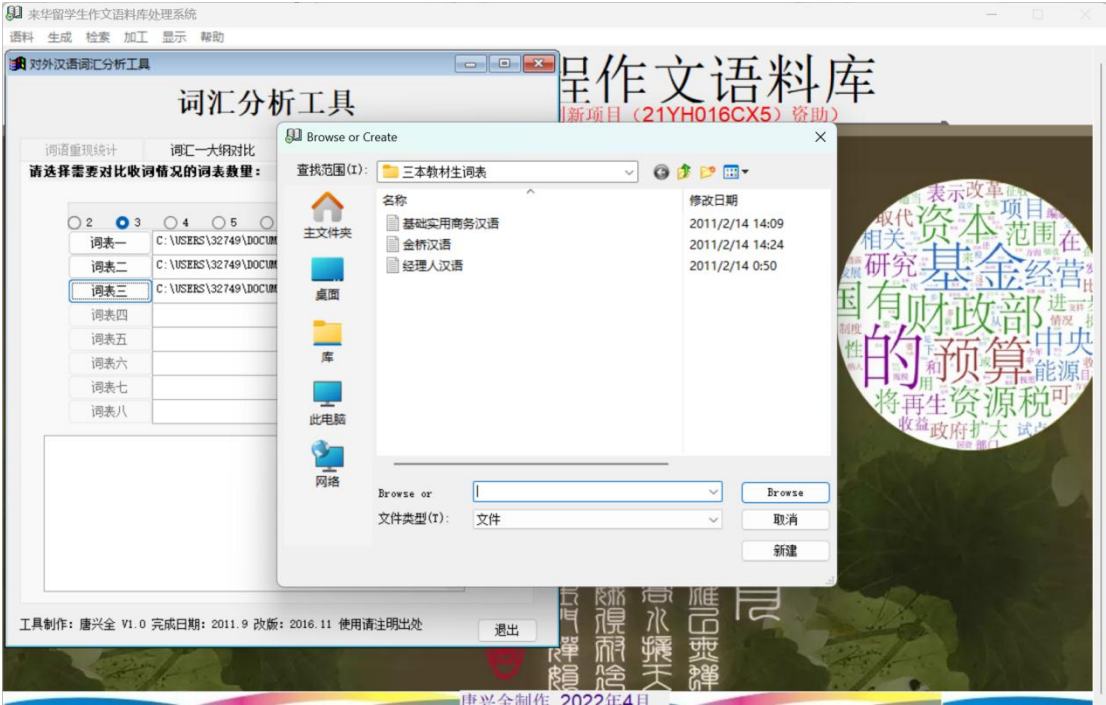


图 54 词汇分析工具操作界面 5

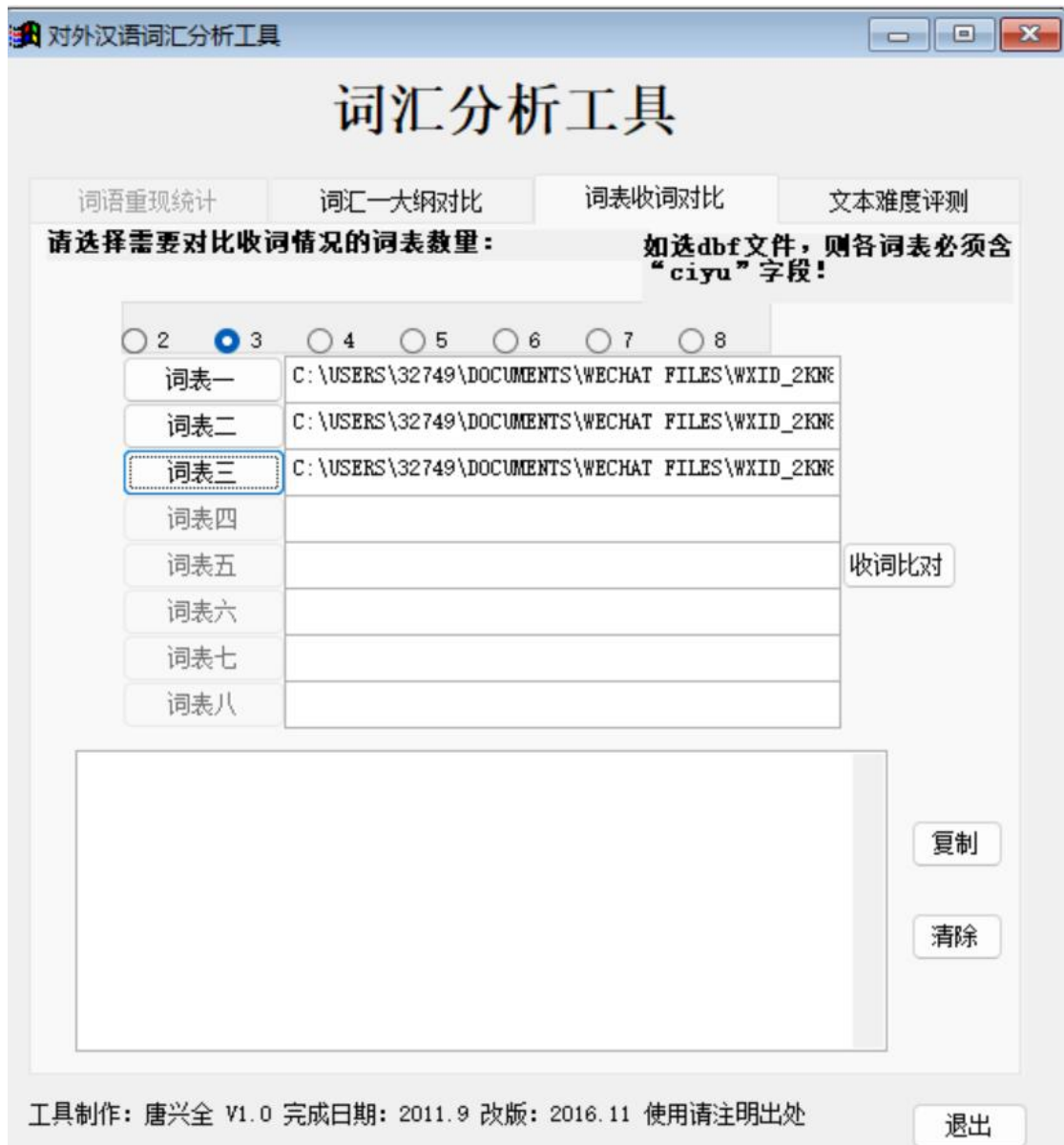


图 55 词汇分析工具操作界面 6

③选择保存结果的文件位置

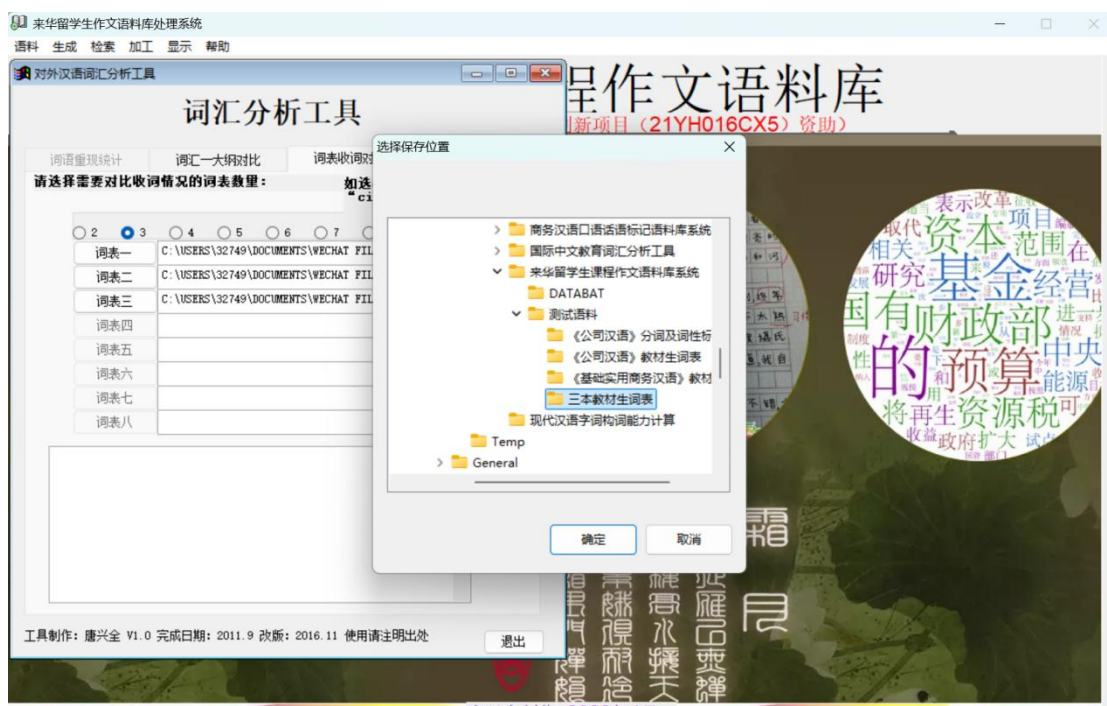


图 56 词汇分析工具操作界面 7

④词表的结果显示在下方窗口。

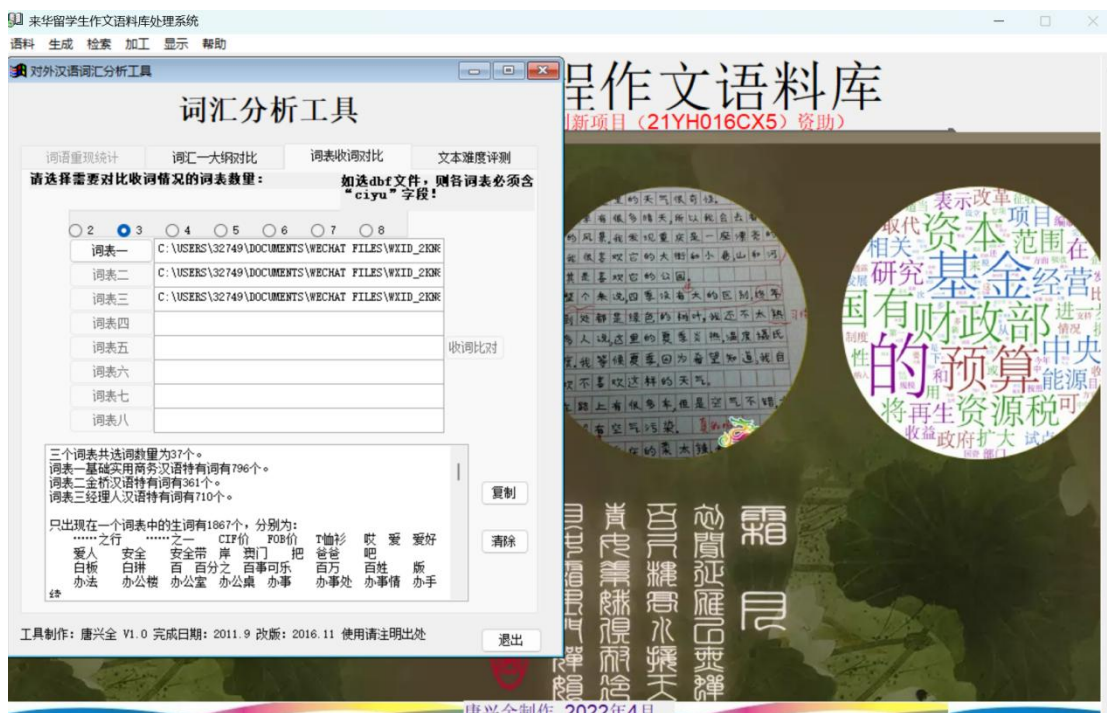


图 57 词汇分析工具操作界面 8

#### 3.2.5.3.4 文本难度测评

根据词语在等级大纲中的分布,判断语料中的词语难度水平,从而推测整篇文本的难度,

### 3.2.5.4 搭配抽取

对词语的前后搭配进行抽取，可查看与某一词的常用搭配词汇。

### 3.2.6 语料显示

将语料文件转化为 XML 文件。



图 58 语料显示选项卡

### 3.2.7 帮助

包含该平台的测试语料和使用说明。