



北京语言大学
BEIJING LANGUAGE AND CULTURE UNIVERSITY

CTAP 汉语文本复杂度分析平台 用户手册

创建团队：BLCU-ICALL

参与编者：朱君辉 崔悦 文彬 徐萌 宗绪泉 杨麟儿 杨尔弘

2023 年 12 月版

（自 2021 年 12 月起编制）

北京语言大学国家语言资源监测与研究平面媒体中心

目录

- 1 用户登录与注册模块 3
- 2 CTAP 平台主界面 6
- 3 语料库管理功能 7
 - 3.1 新建语料库 8
 - 3.2 管理语料库 14
- 4 特征提取 15
 - 4.1 创建特征集 16
 - 4.2 选择语言特征 18
- 5 特征分析功能 22
 - 5.1 新建分析项目 23
 - 5.2 进行特征分析 24
- 6 结果可视化功能 26

1 用户登录与注册模块

访问 CTAP—中文文本复杂度分析平台首页，用户可以通过首页中间的登录框登录，登录成功后，通过输入邮箱和密码对平台并点击【登录/login】进入平台，如图 1、图 2 所示：

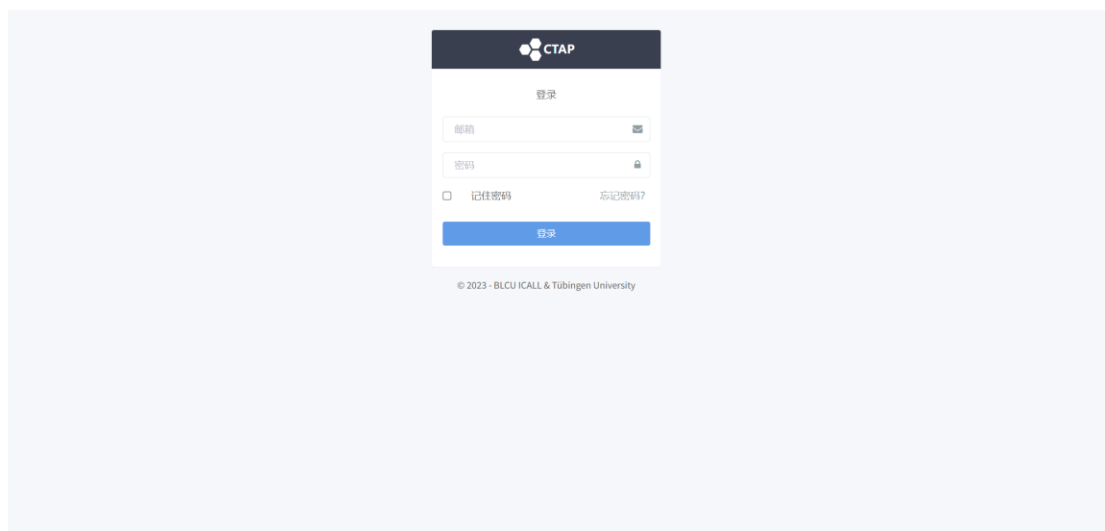


图 1 平台首页登录页面-中文版

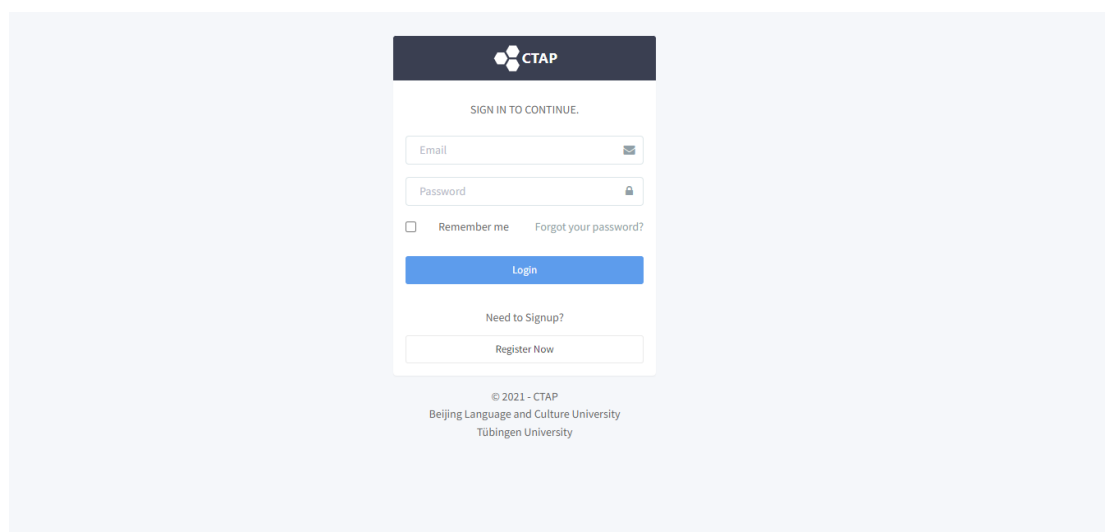
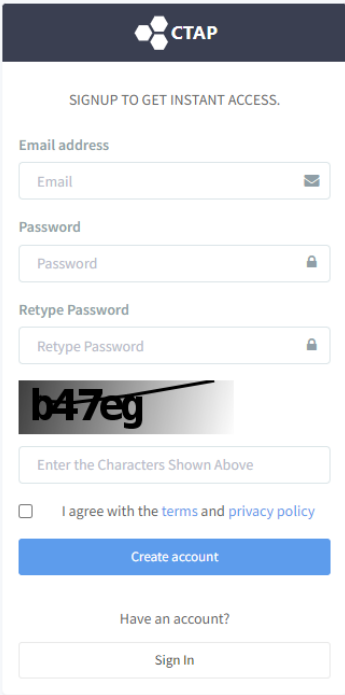


图 2 平台首页登录页面-英文版

用户登录汉语文本复杂度分析平台后，可以管理与上传语料库进行特征的计算，在此平台上进行中文或跨语言文本复杂度分析研究。该平台采用模块化的非结构化信息管理框架（UIMA），支持将不同语

言分析模块添加到现有的系统结构中。未登录的用户不能使用平台，用户可以通过点击【Register Now】，进入注册页面如图 3 所示，完成注册并且登录。（暂停使用，请联系管理员）



The image shows a user registration form for the CTAP system. The form is titled "SIGNUP TO GET INSTANT ACCESS." and includes fields for "Email address", "Password", and "Retype Password". Below these fields is a CAPTCHA image showing the characters "b47eg". A checkbox for "I agree with the terms and privacy policy" is present, followed by a blue "Create account" button. At the bottom, there is a link for "Have an account?" with a "Sign In" button. The footer indicates "© 2016 - CTAP Tübingen University".

图 3 用户注册

用户忘记密码无法登录，可以点击首页【Forgot your password?】进入密码找回页面如图 4 所示，用户可以通过邮箱验证将自己的密码重置。（暂停使用，请联系管理员）

CTAP

PASSWORD RESET

Fill with your mail to receive instructions on how to reset your password.

Email address

Reset

© 2016 - [CTAP](#)
Tübingen University

图 4 找回密码

2 CTAP 平台主界面

CTAP 平台支持中英文切换，中英文版主界面如图 5、6 所示。

左侧导航栏分别为语料库管理、特征选择、特征分析与结果可视化四大功能。

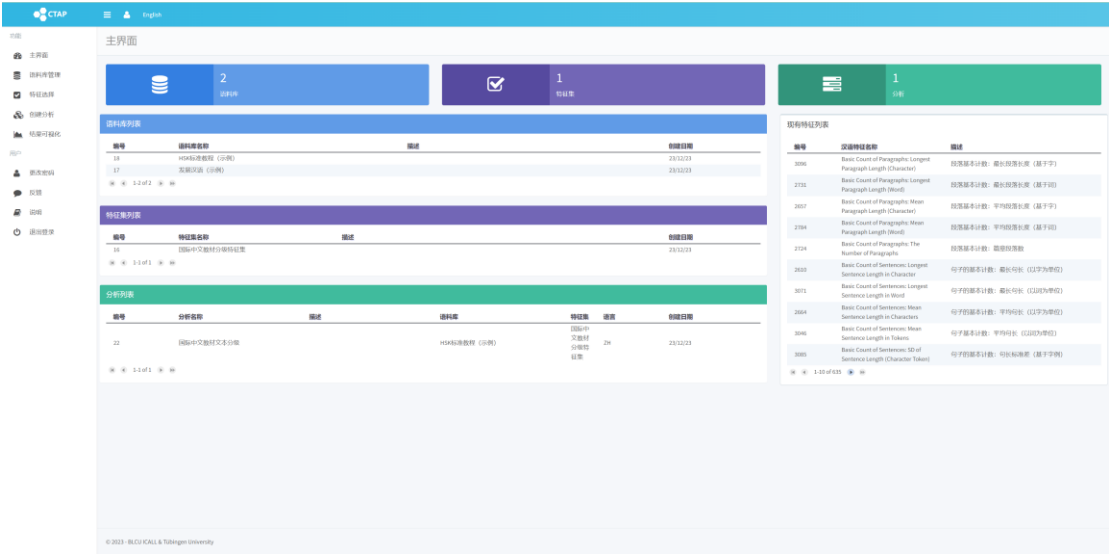


图 5 CTAP 主界面-中文版

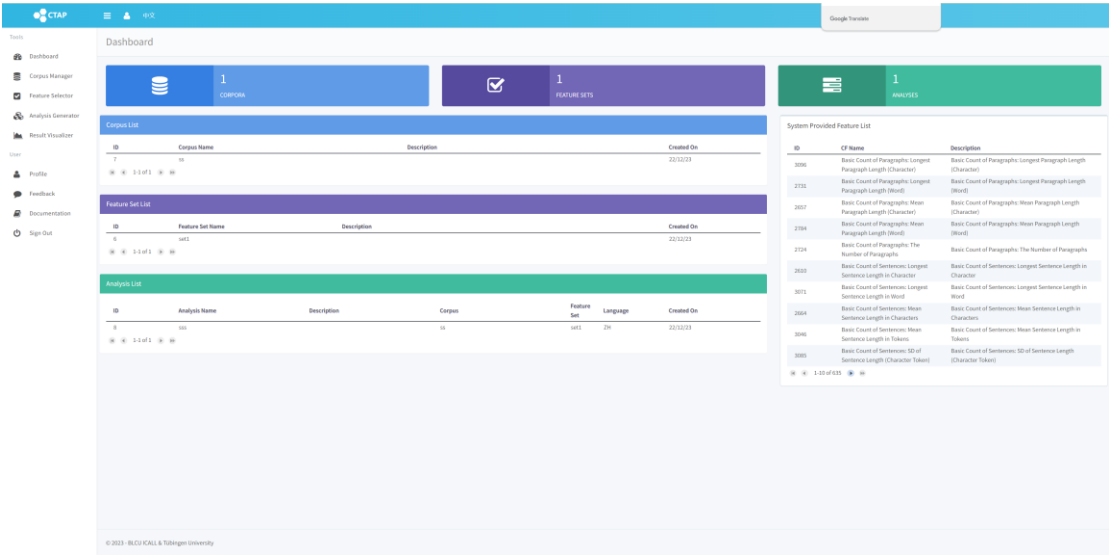


图 6 CTAP 主界面-英文版

3 语料库管理功能

语料库管理支持用户上传多个文件，并在平台上建立属于自己的语料库。用户可以通过创建文件夹来对语料库进行分类管理，然后创建语料库来储存上传的文本，可通过添加标签来对具体的文本做标记。如图 7、图 8 所示，

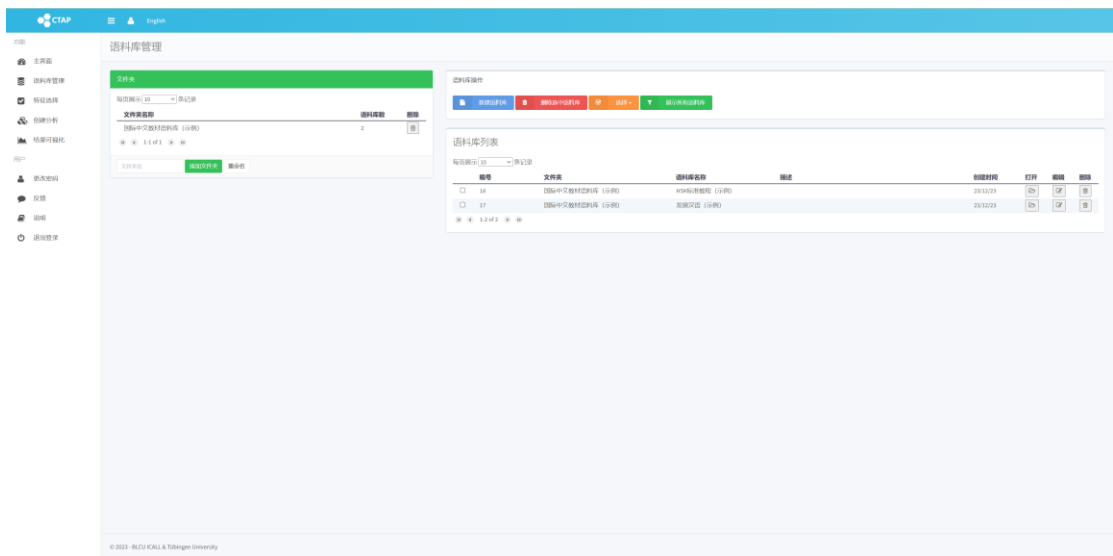


图 7 语料库管理功能-中文版

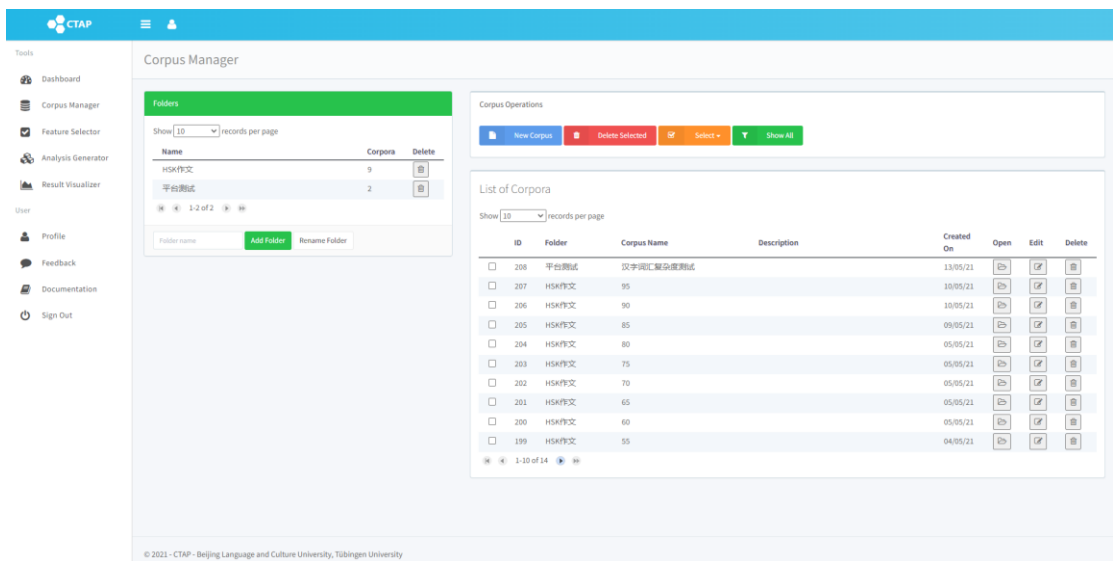


图 8 语料库管理功能-英文版

3.1 新建语料库

在【文件夹名】/【Folder Name】处输入要建的新文件夹的名称，
点击【添加文件夹】/【Add Folder】，显示【文件夹创建成功】/【New folder created】，即新建语料库成功，如图 9、图 10 所示。

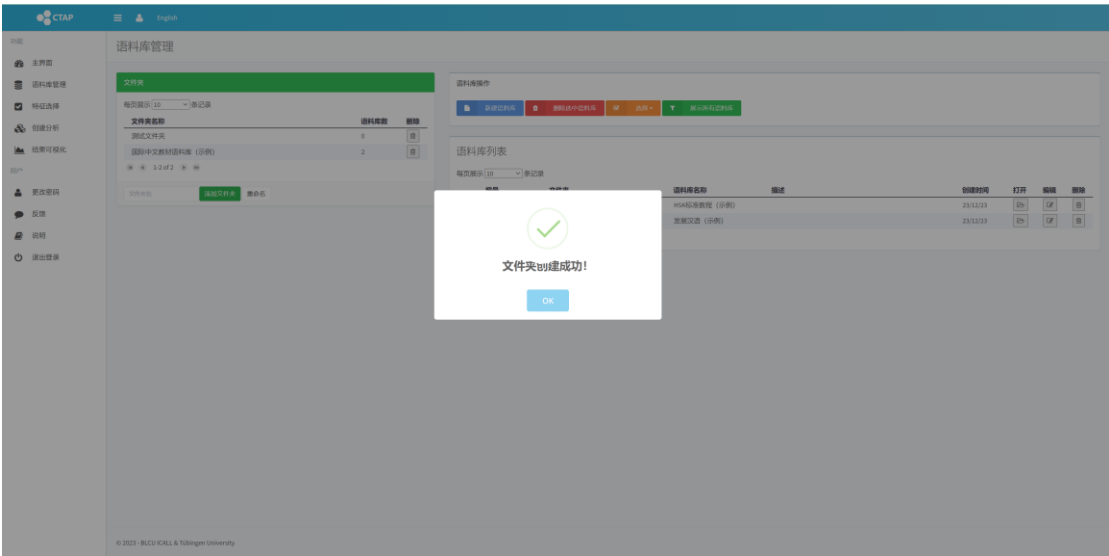


图 9 新建语料库功能-中文版

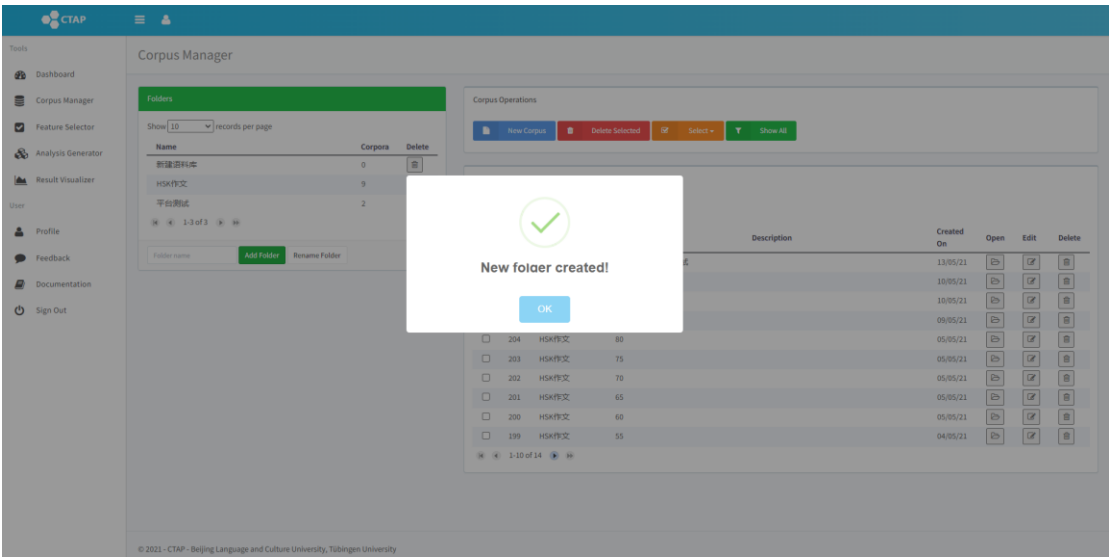


图 10 新建语料库功能-英文版

点击新建的语料库名称，页面右侧是新建语料库下所有子语料库的列表，如图 11、图 12 所示。

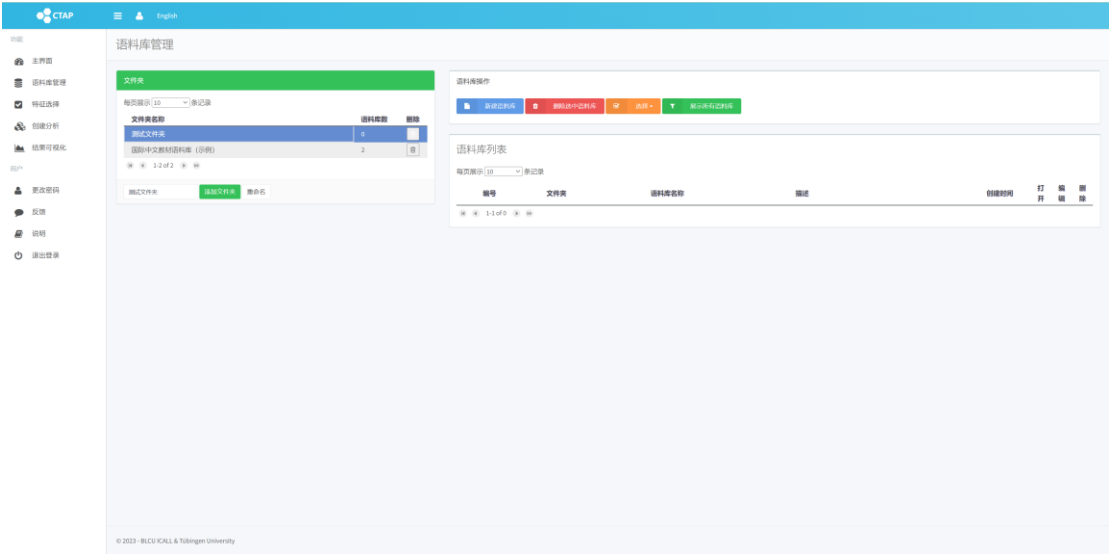


图 11 管理子语料库功能-中文版

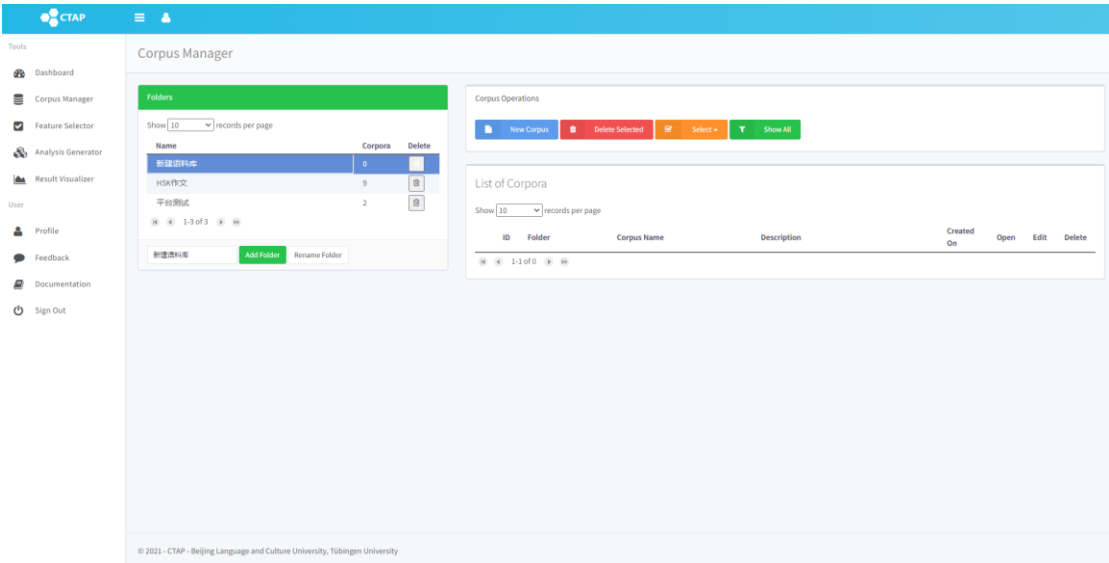


图 12 管理子语料库功能-英文版

点击【新建语料库】/【New Corpus】新建子语料库，会显示需要填写语料库详情，包括位于哪个文件及、名称以及描述。如图 13、图 14 所示。

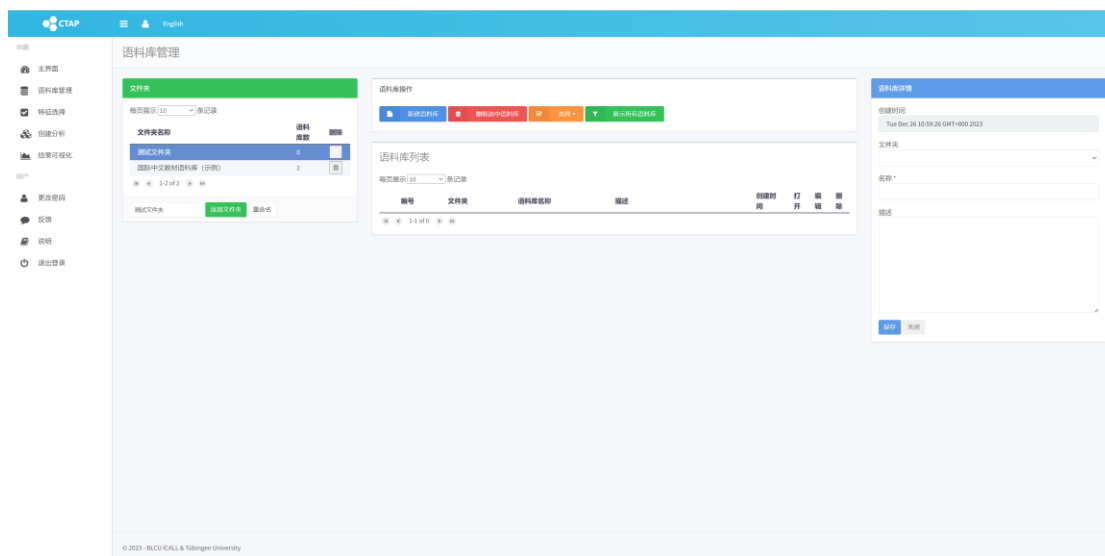


图 13 新建子语料库-中文版

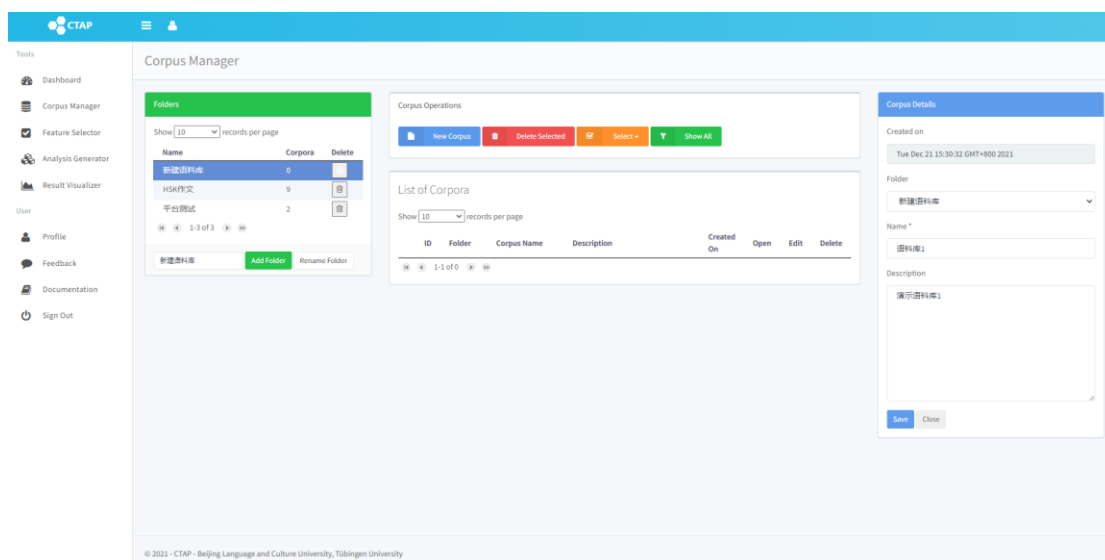


图 14 新建子语料库-英文版

点击【保存】/【Save】，返回语料库管理的页面，语料库已新建成功，如图 15、图 16：

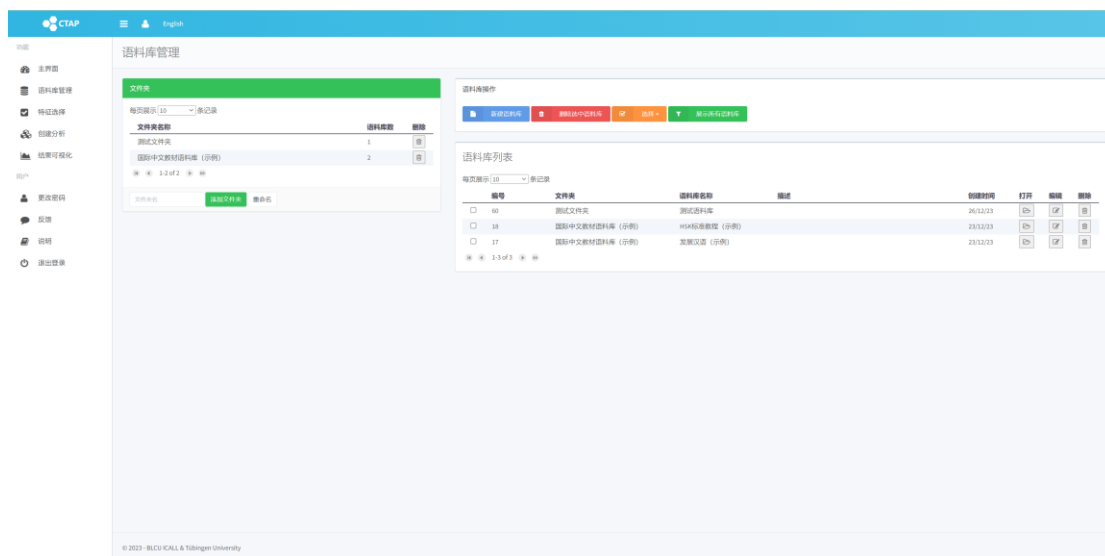


图 15 查看语料库-中文版

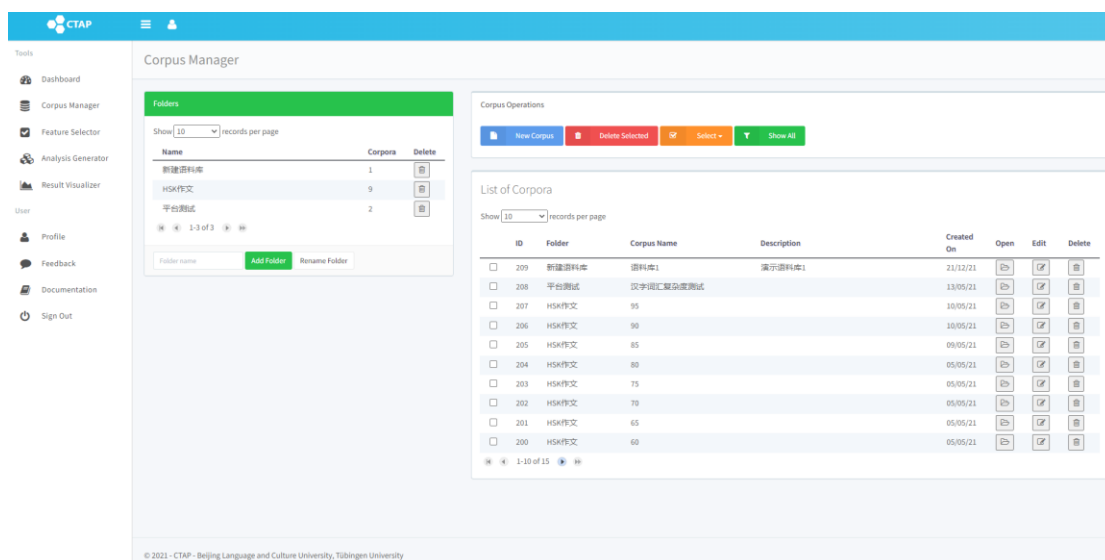


图 16 查看语料库-英文版

查看具体的内容，点击【打开】/【open】，如图 17、图 18：

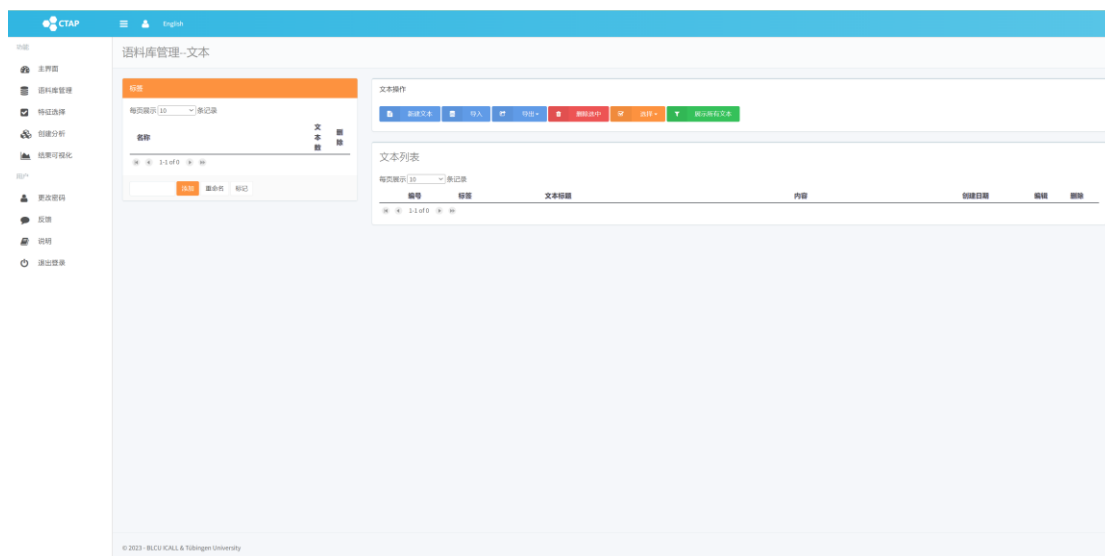


图 17 某一具体语料库-中文版

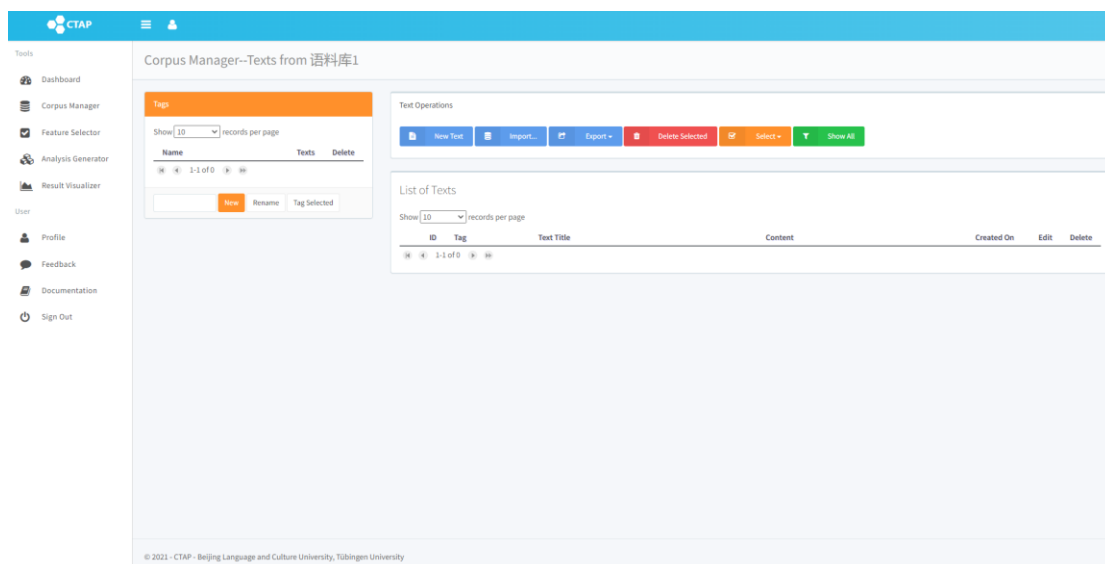


图 18 某一具体语料库-英文版

在【文本操作】/【Text Operations】栏中，可以选择新建文本内容，可以选择上传本地语料、导出语料或者删除语料。

新建文本内容，点击【新建文本】/【New Text】，如图 19、图 20 所示：

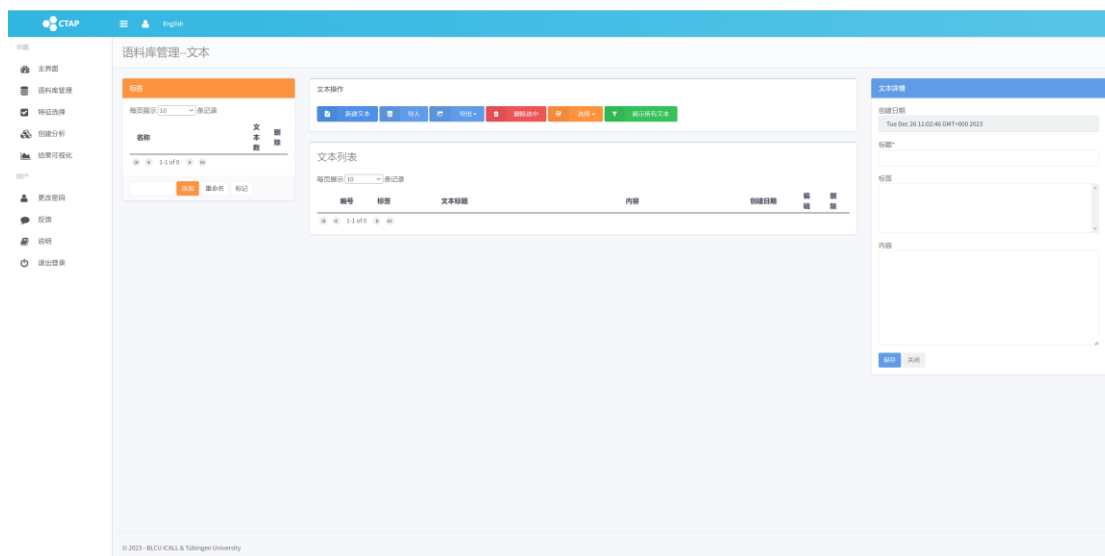


图 19 新建文本内容-中文版

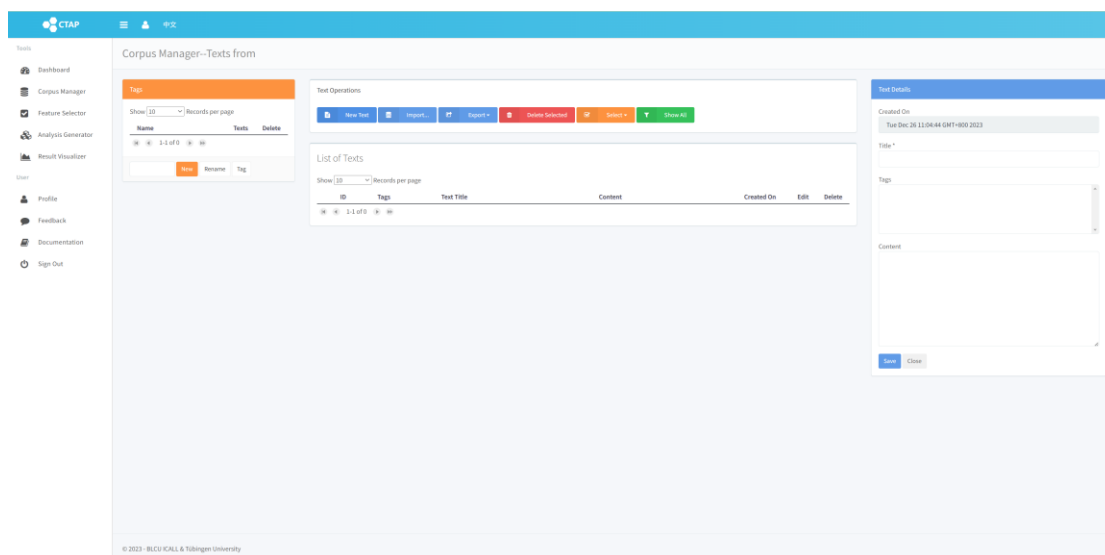


图 20 新建文本内容-英文版

上传本地语料，点击【导入】/【Import Texts】，点击【选择文件】/【Add files】，从本地文件夹中选择需要上传的 txt 文件，上传。如图 21、图 22 所示：

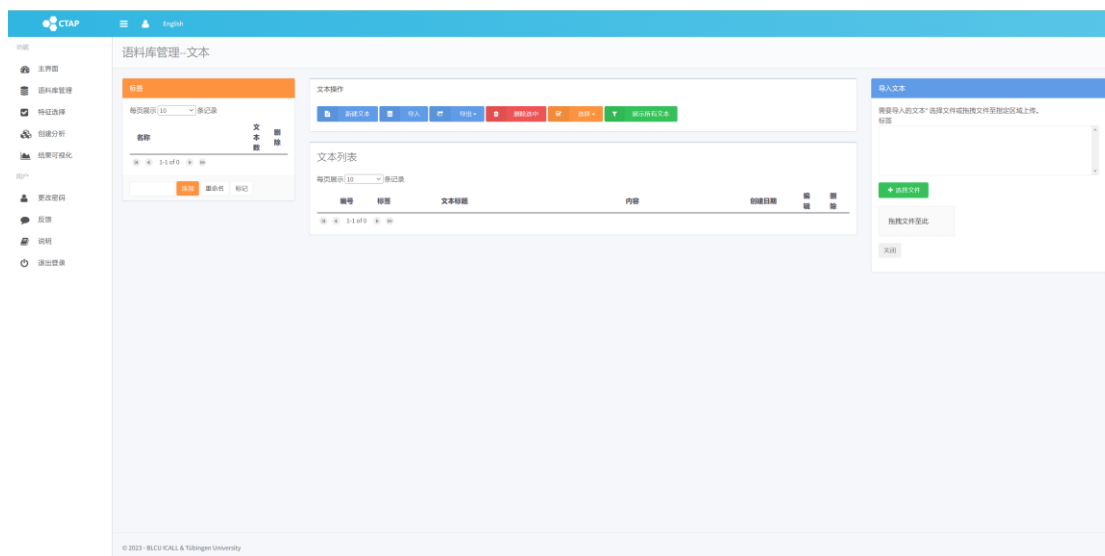


图 21 导入本地语料-中文版

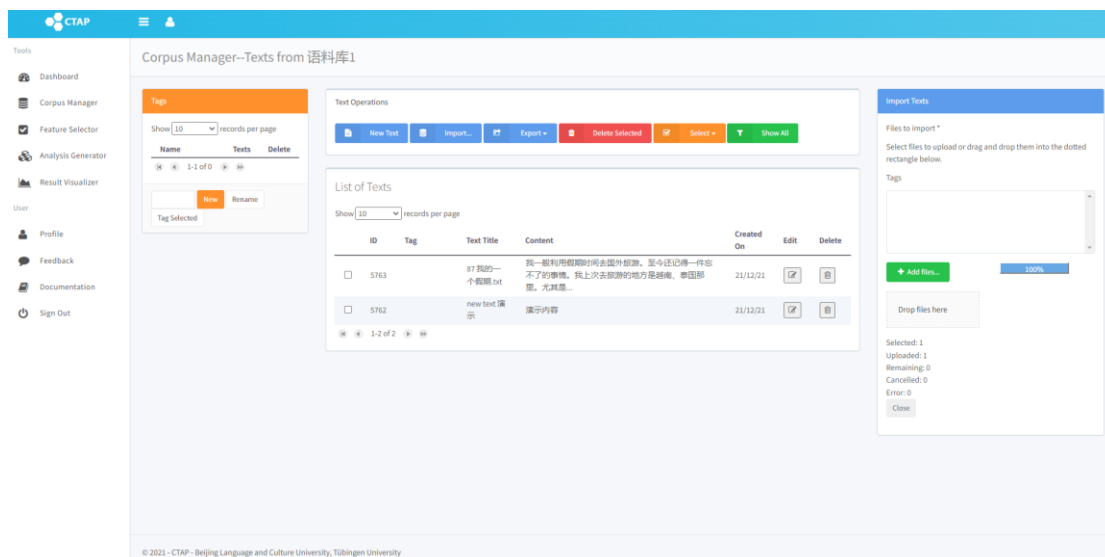


图 22 导入本地语料-英文版

3.2 管理语料库

实现语料库的管理本平台的突出功能，语料库集中可以存放大量用作不同途径的语料库。语料库列表里面显示了每个语料库的编号、名称、标签、创建时间，如图 23、图 24 所示：

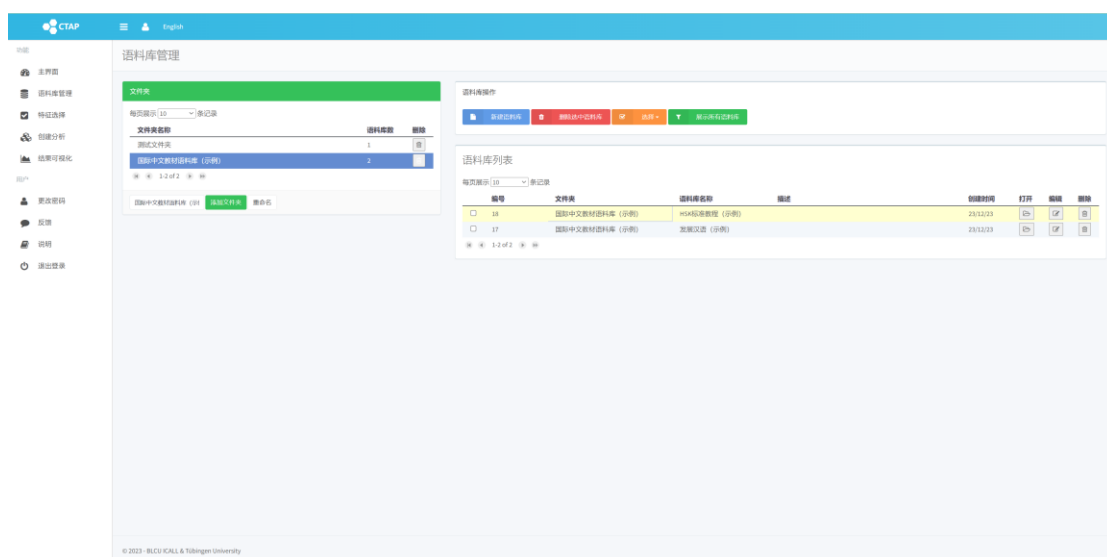


图 23 语料库管理-中文版

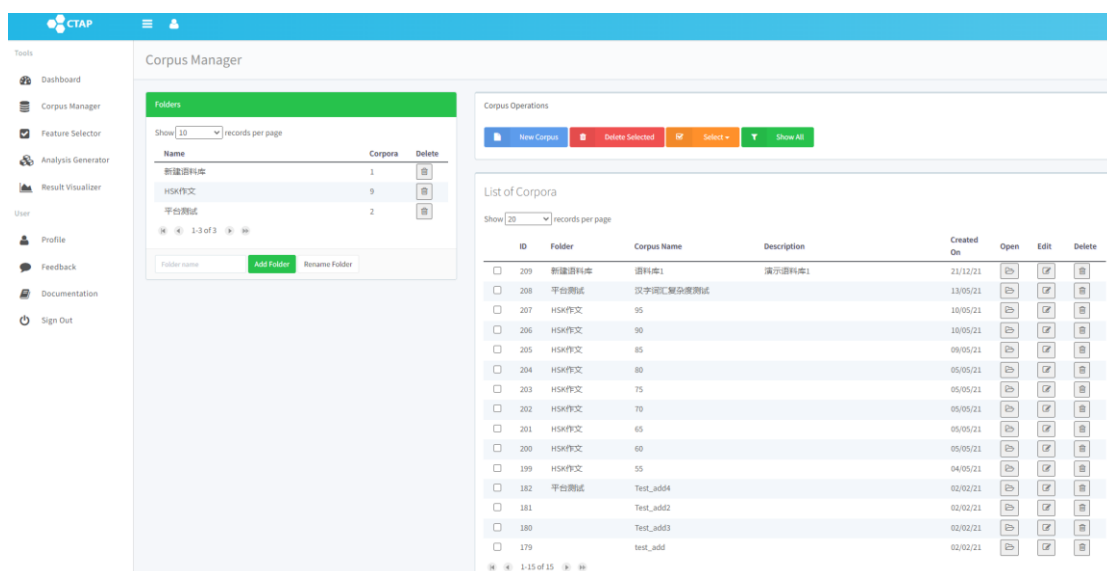


图 24 语料库管理-英文版

4 特征提取

不同于现有的文本复杂度分析工具，中文 CTAP 允许用户根据自己的研究目的创建特征集。首先，用户需要创建一个空的特征集，然后在现有的特征列表里选择复杂度特征加入到自己的特征集。在选择特征时，用户可通过选择具体的语言来提取特征，也可在搜索框输入所

想要提取的特征。

点击【特征选择】/【Feature Selector】，进入特征选择页面，如图 25、图 26；

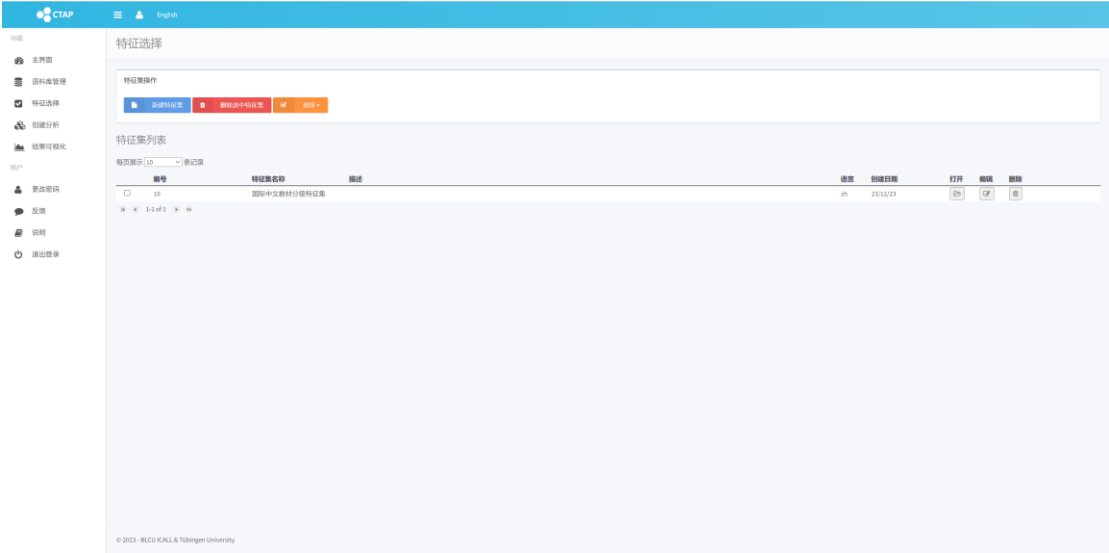


图 25 特征选择界面-中文版

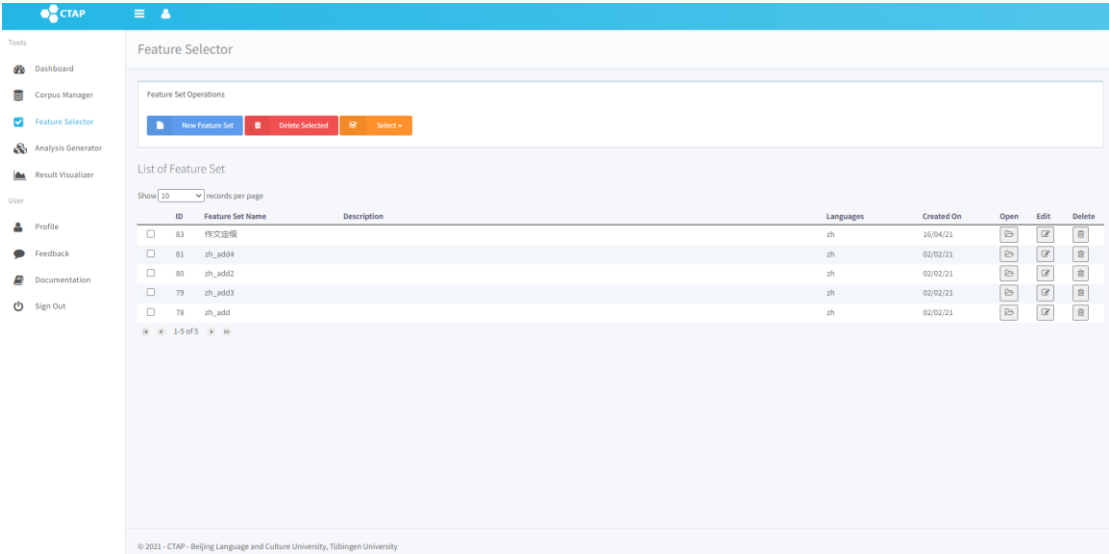


图 26 特征选择界面-英文版

4.1 创建特征集

特征集是进行文本复杂度分析的重要组成部分，本系统的特征在

英语、意大利语特征集的基础上，增加了汉语特有的特征，针对性地进行汉语文本复杂度分析。本系统可以同时管理多个特征集，用户可以根据研究的需求创建不同的特征集。

进入页面后点击【新建特征集】/【New Feature Set】，可以创建新的特征集，填好基本信息之后，点击【保存】/【Save】便创建成功了。

如图 27、图 28：

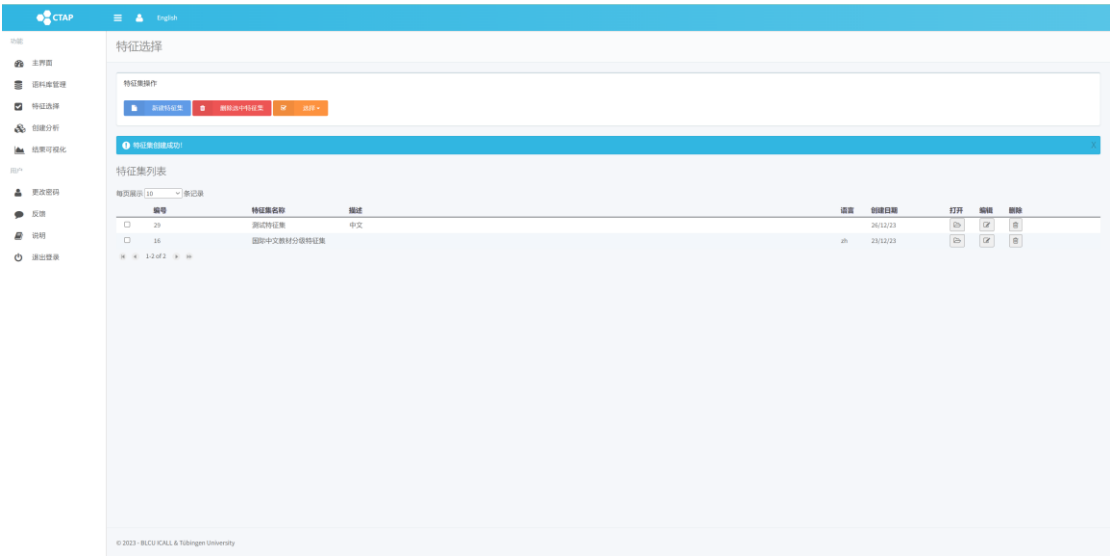


图 27 创建特征集-中文版

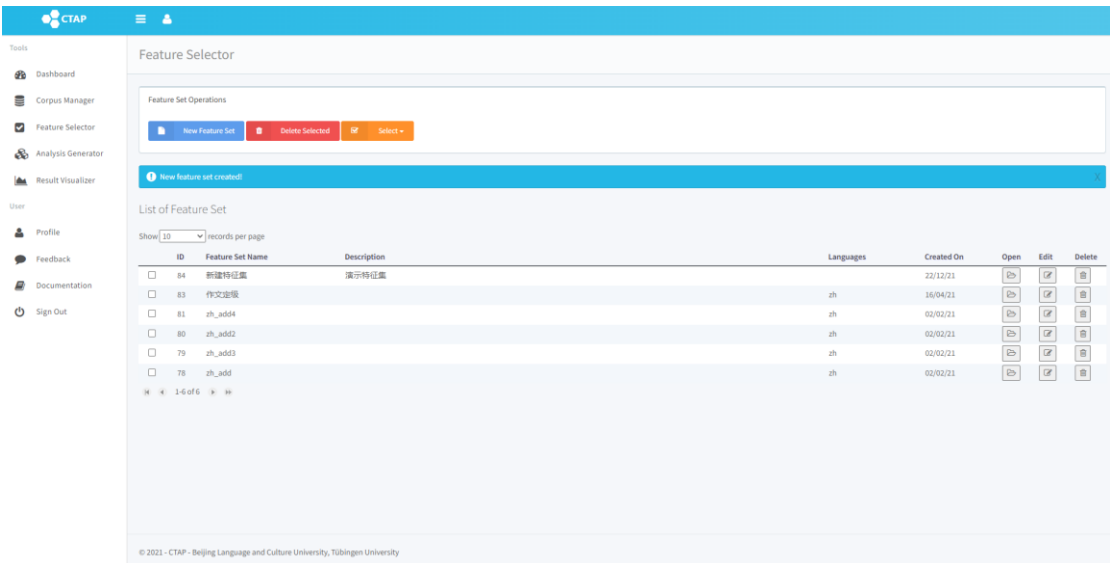


图 28 创建特征集-英文版

4.2 选择语言特征

点击【打开】/【open】进入新创建好的特征集， 左侧为用户创建的特征集，目前是空白；右侧【可选特征】/【Available Features】为待添加的特征。如图 29、图 30 所示。

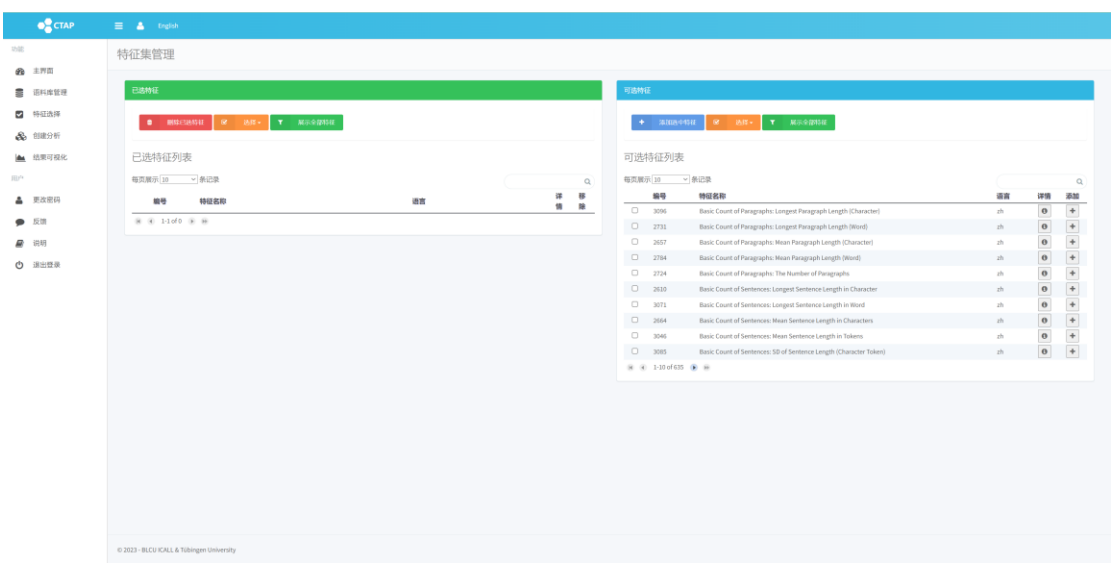


图 29 进入新建特征集页面-中文版

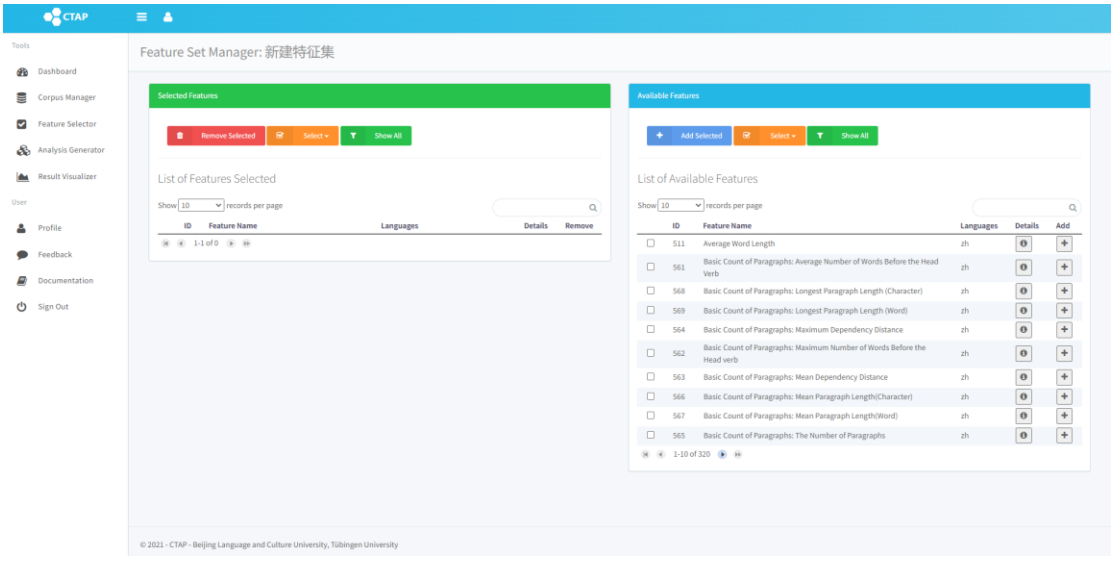


图 30 进入新建特征集页面-英文版

点击【选择】/【Selected】可以根据语言选择特征；点击【展示全部特征】/【Show All】查看所有特征。

添加新的特征到新建特征集中有三种方式：

1、每一行特征后面都有两个按钮，【zh】表示语言为中文，点击【详情】/【Details】可以查看该特征的详细信息， 点击【+】可以直接逐个添加到左侧的新建特征集中，如图 31、图 32；

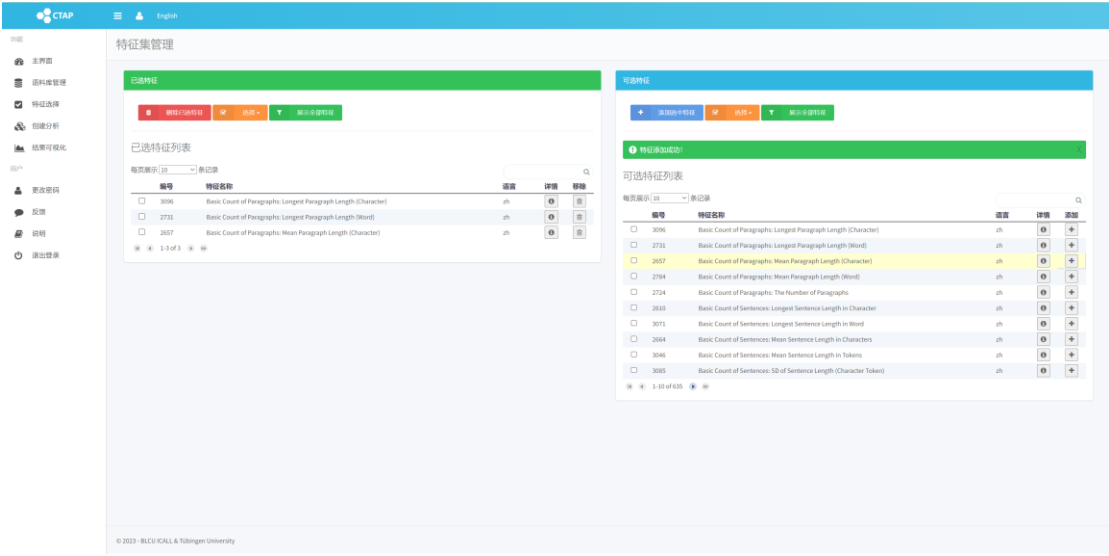


图 31 逐个添加特征-中文版

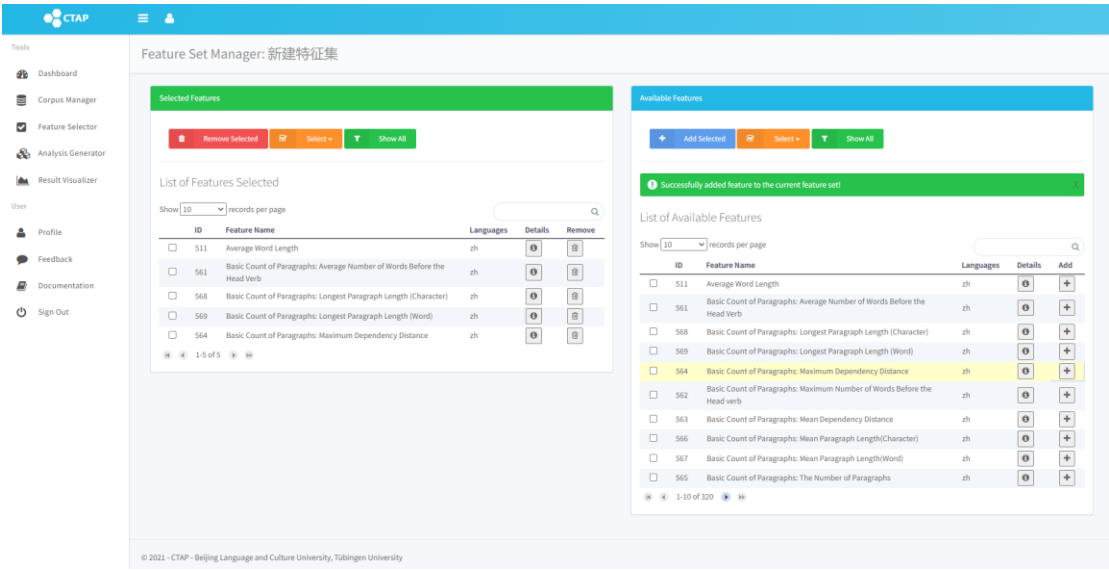


图 32 逐个添加特征-英文版

2、点击【选择】/【Selected】，选择【中文】/【Chinese】，再点击【添加选中特征】/【Add Selected】，即可把所有的中文特征添加到新建特征集中，如图 33、图 34 所示；

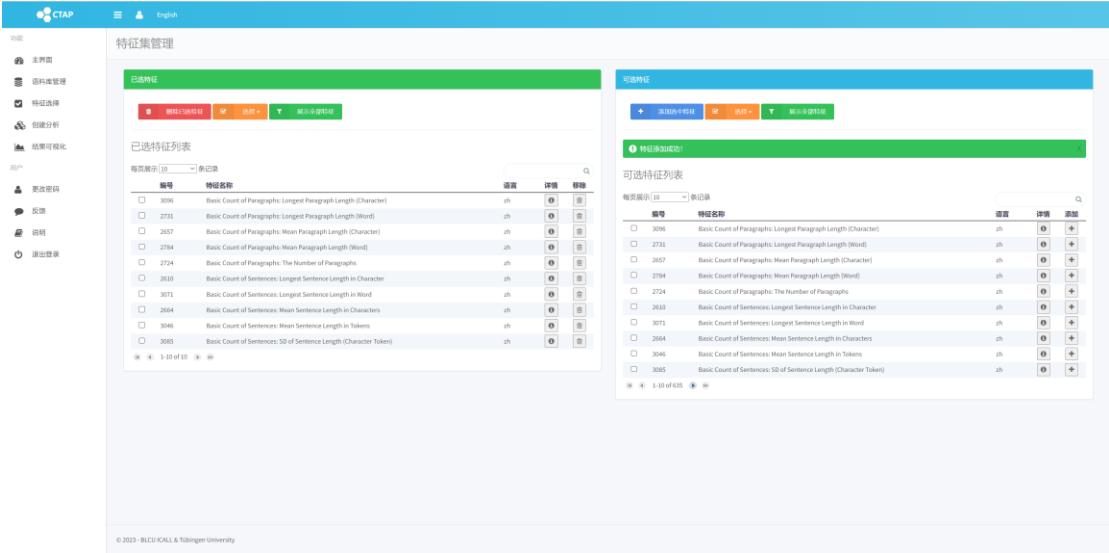


图 33 批量添加特征-中文版

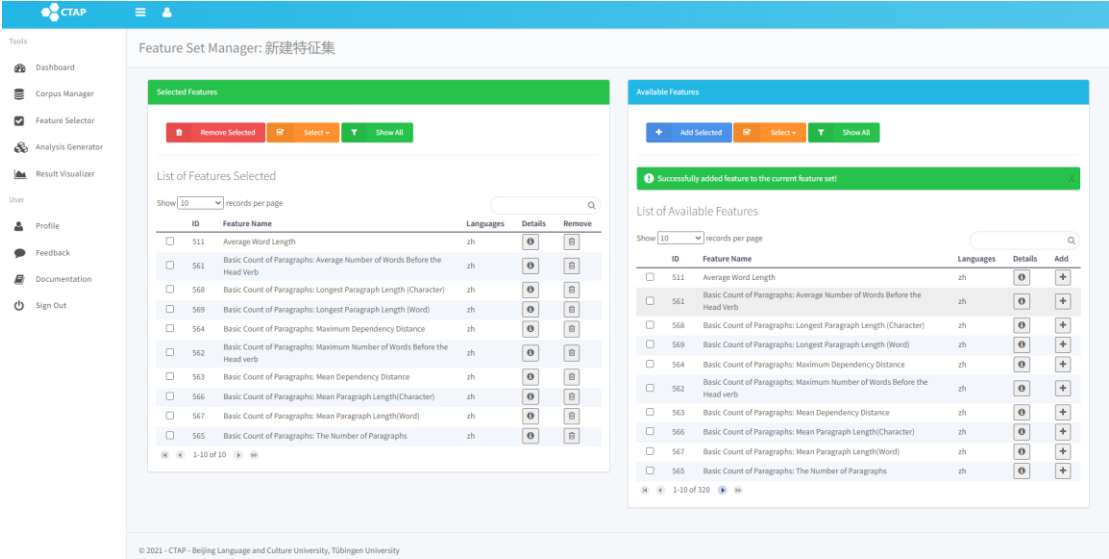


图 34 批量添加特征-英文版

在中文版中，点击特征后的详情，可看到该特征的中文描述，如图 35。

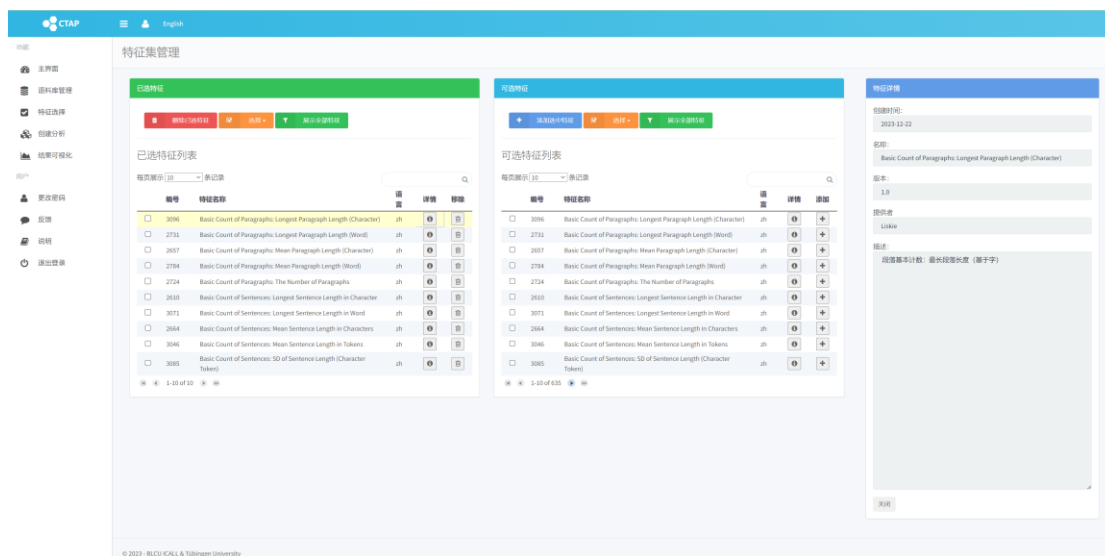


图 35 特征中文描述-中文版

3、在搜索框中直接搜索需要的特征，点击【+】添加到新建特征集中，如图 19 所示。

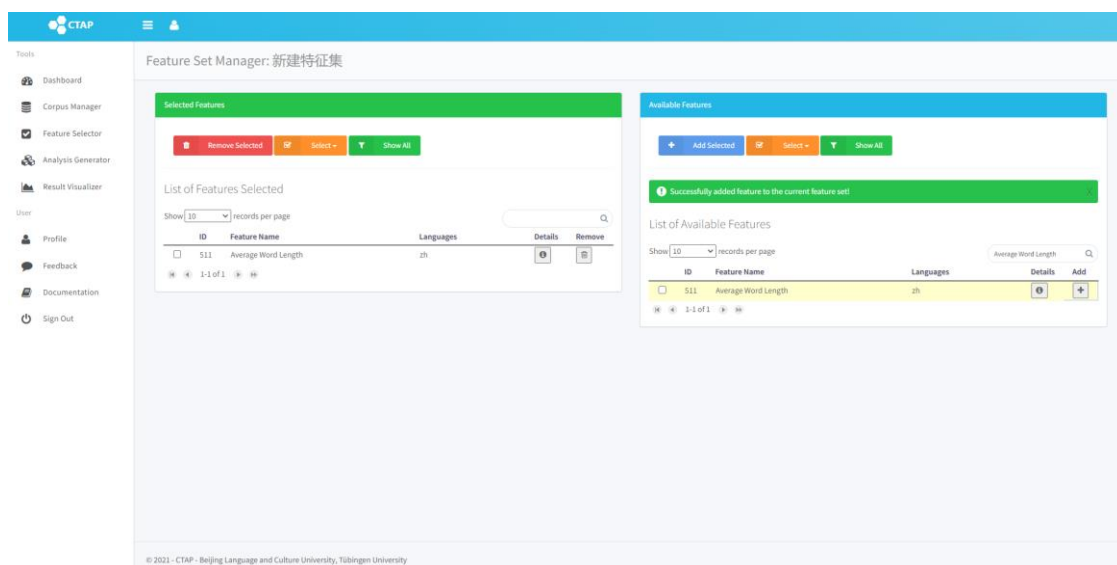


图 36 搜索添加特征

5 特征分析功能

在选定语料和特征集后，可以通过 Analysis Generator 生成分析结果并保存，也支持将分析结果以 CSV 格式下载到本地，进行进一步处理和分析。

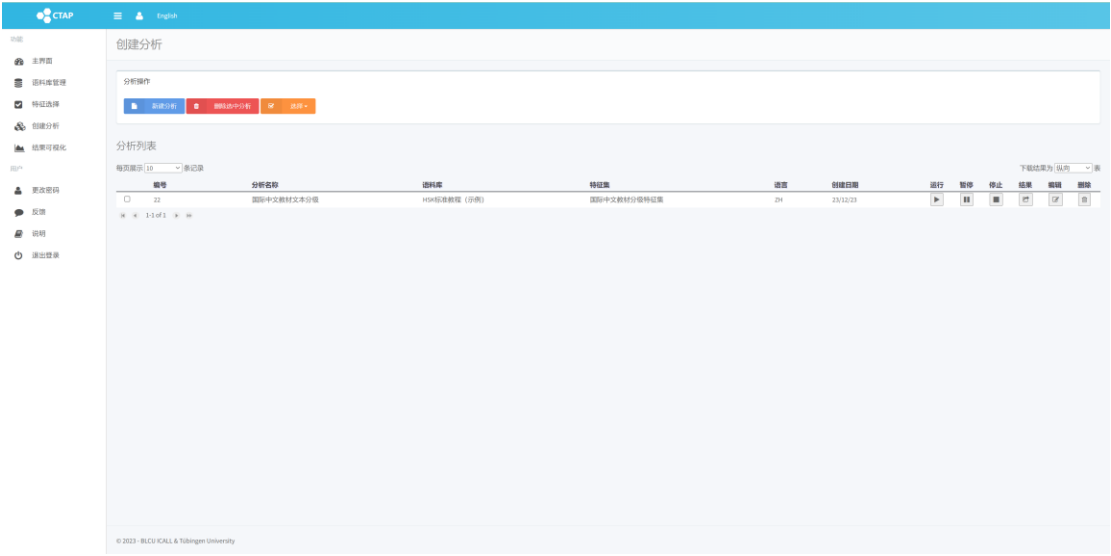


图 37 分析管理页面-中文版

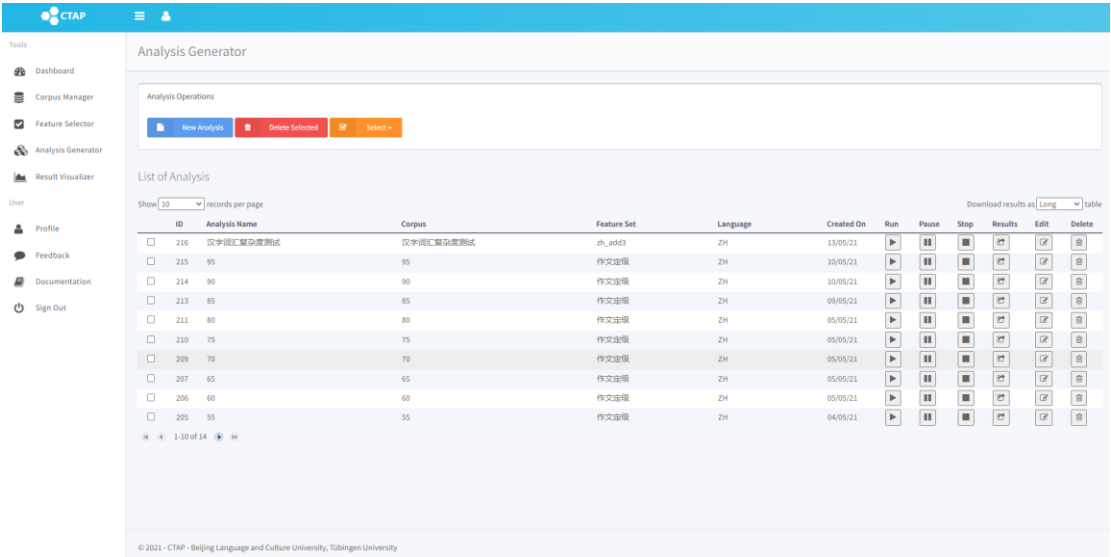


图 38 分析管理页面-英文版

5.1 新建分析项目

点击【新建分析】/【New Analysis】，新建分析，在右侧填写详细信息，【名称】/【Name】处填写分析项目的名称，【语言】/【Language】选择【中文】/【Chinese】，【选择一个语料库】/【Select a Corpus】选择要分析的语料库，【选择一个特征集】/【Select a Feature Set】选择要使用的特征集，选择完毕后点击【保存】/【Save】，新建分析项目成功。如图 39、图 40 所示。

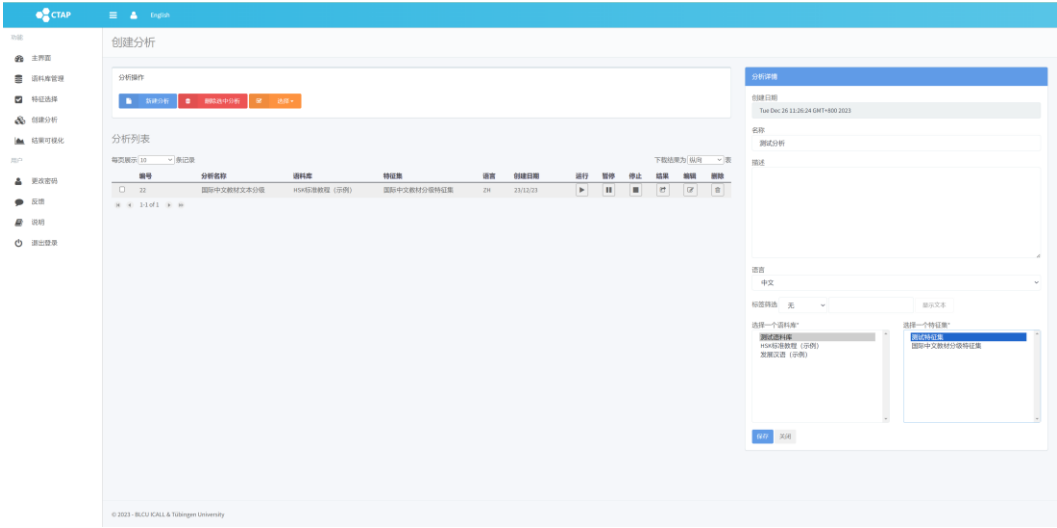


图 39 新建分析项目-中文版

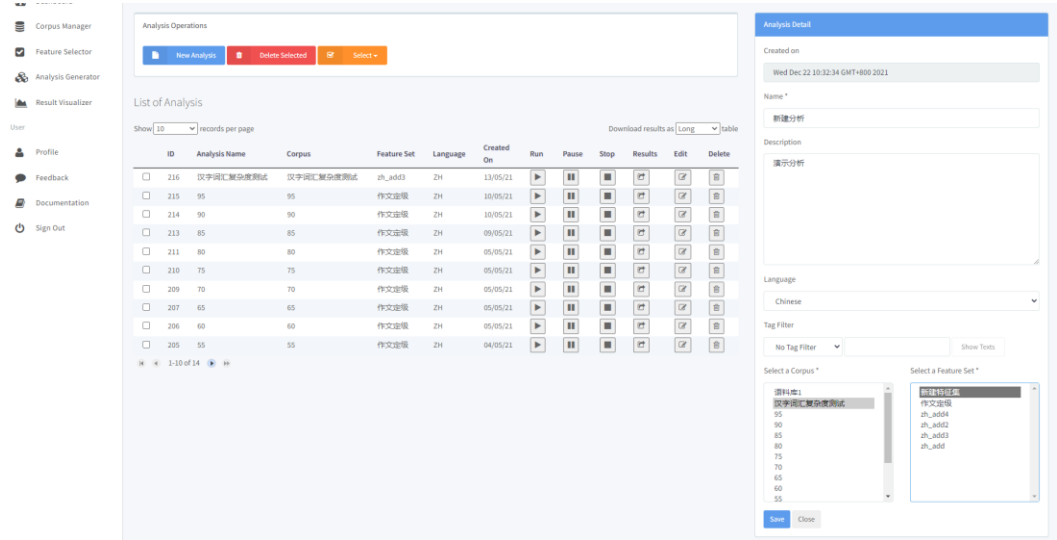


图 40 新建分析项目-英文版

5.2 进行特征分析

分析项目新建成功之后，点击后方的【运行】/【Run】即可开始分析，如图 41、图 42 所示。

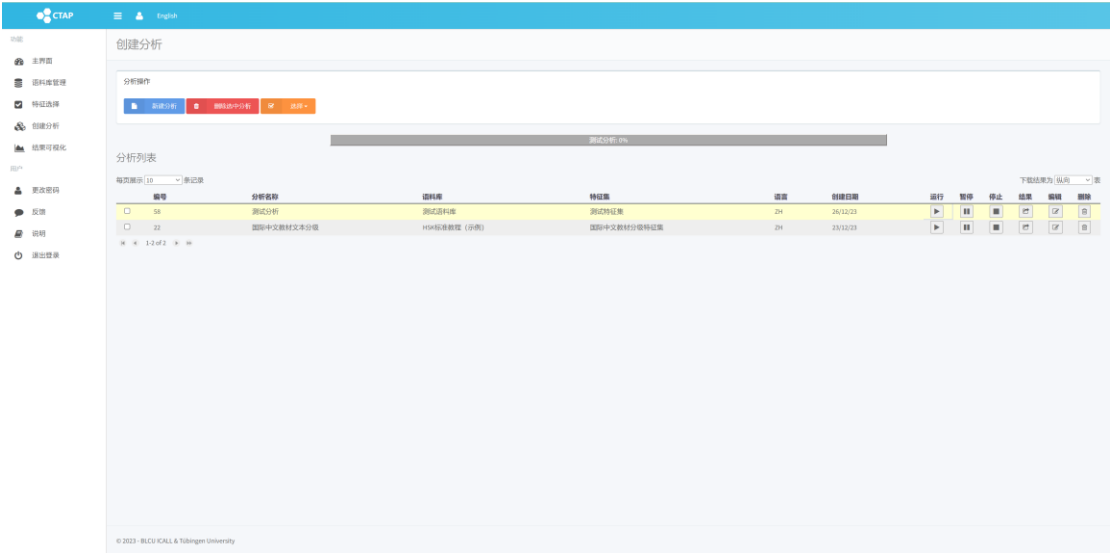


图 41 开始分析-中文版

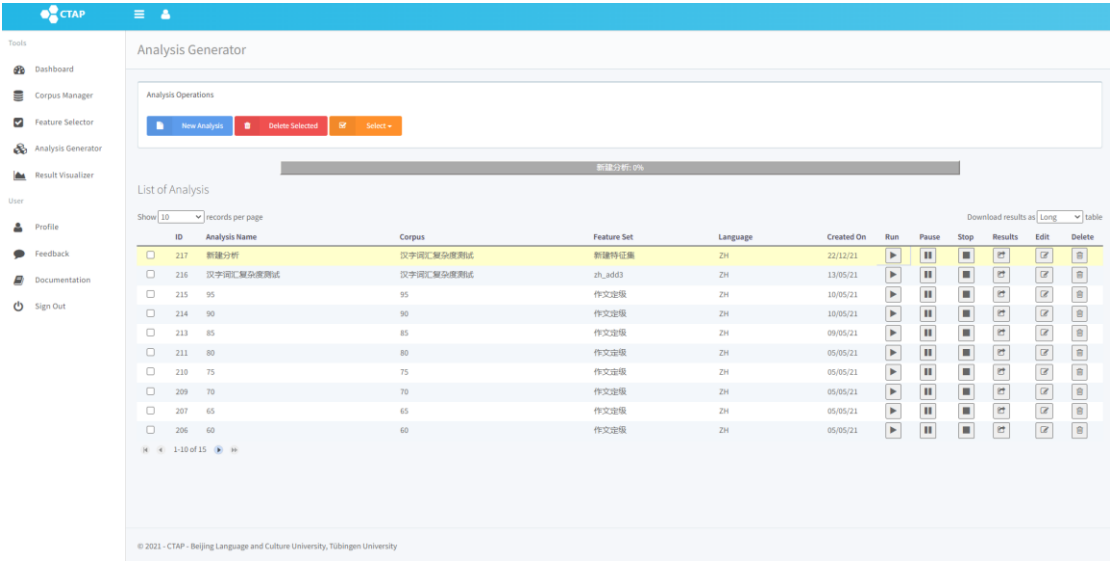


图 42 开始分析-英文版

点击【暂停】/【Pause】暂停分析，点击【停止】/【Stop】停止分析。分析进度条到 100%即分析结束，如图 43、图 44 所示。

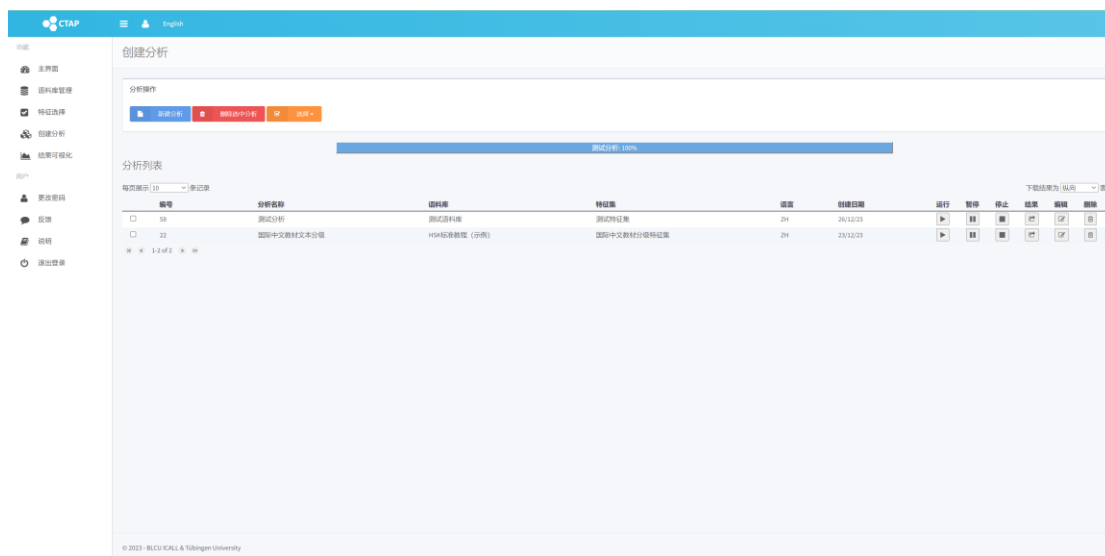


图 43 分析成功-中文版

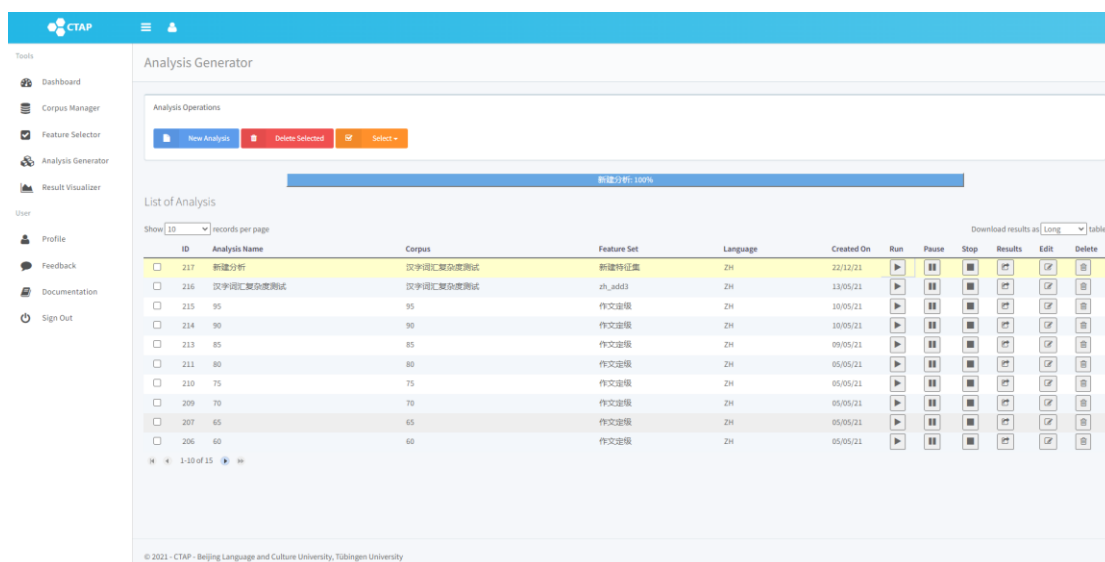


图 44 分析成功-英文版

点击【结果】/【Results】按钮，下载导出分析结果，如图 45。

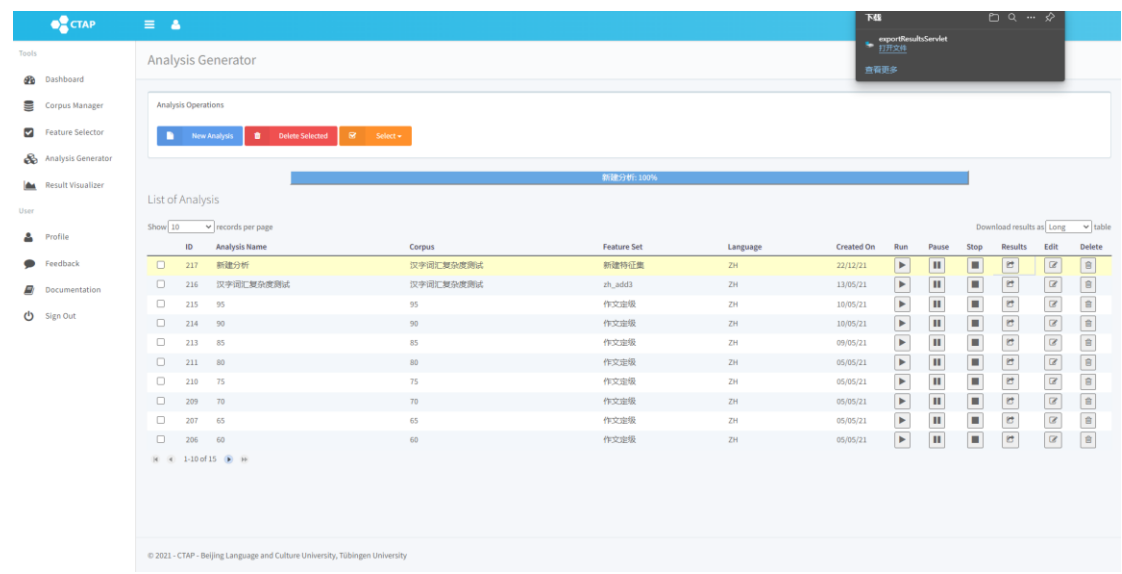


图 45 导出结果

下载成功后打开文件，结果如图 47 所示。

1	Text_id	Tags	Text_Title	Feature_id	Feature_Name	Value
2	5761	test	511	Average Word Length	1.4285714285714286	
3	5761	test	561	Basic Count of Paragraphs: Average Number of Words Before the Head Verb	0.0	
4	5761	test	562	Basic Count of Paragraphs: Maximum Number of Words Before the Head verb	0.0	
5	5761	test	563	Basic Count of Paragraphs: Mean Dependency Distance	2.5714285714285716	
6	5761	test	564	Basic Count of Paragraphs: Maximum Dependency Distance	7.0	
7	5761	test	565	Basic Count of Paragraphs: The Number of Paragraphs	1.0	
8	5761	test	566	Basic Count of Paragraphs: Mean Paragraph Length(Character)	11.0	
9	5761	test	567	Basic Count of Paragraphs: Mean Paragraph Length(Word)	8.0	
10	5761	test	568	Basic Count of Paragraphs: Longest Paragraph Length (Character)	11.0	
11	5761	test	569	Basic Count of Paragraphs: Longest Paragraph Length (Word)	8.0	

图 46 分析结果样例

6 结果可视化功能

可视化模块支持基本的绘图操作，可以帮助用户对分析结果进行可视化展示。

点击【结果可视化】/【Result Visualizer】进入结果可视化页面，如图 47、图 48 所示。

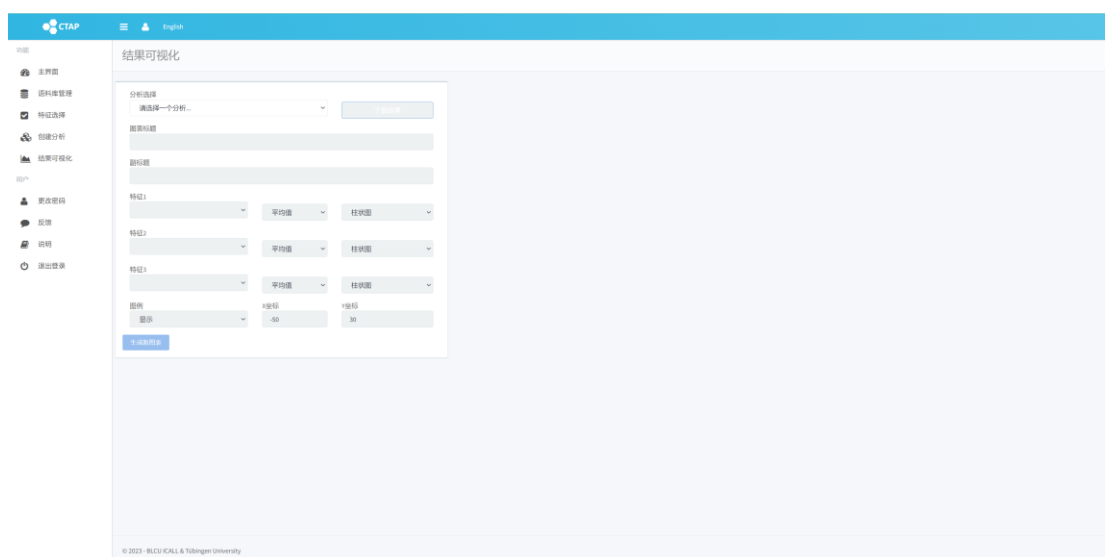


图 47 结果可视化功能-中文版

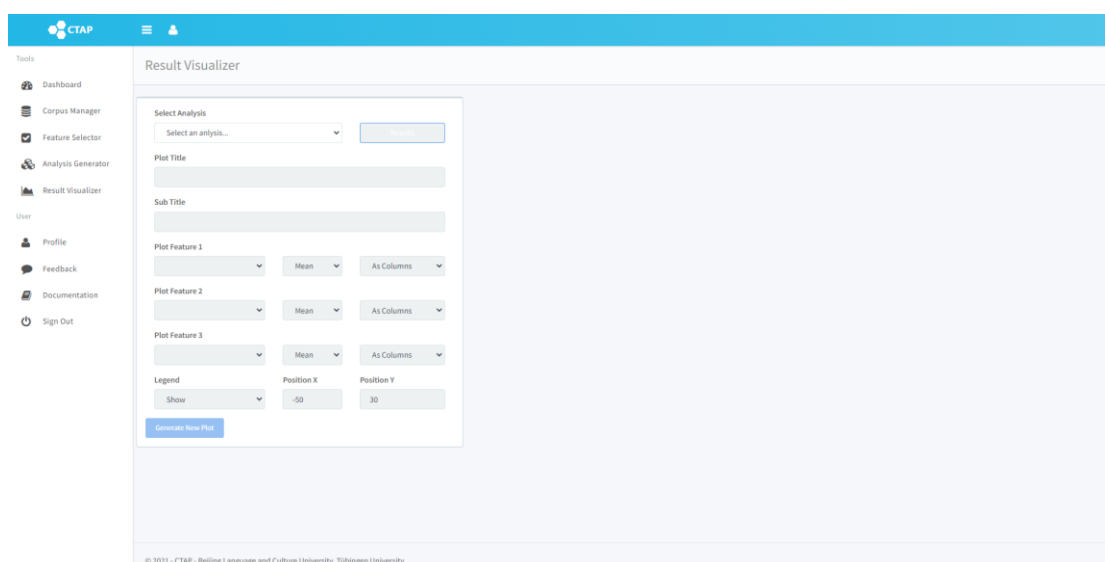


图 48 结果可视化功能-英文版

点击【分析选择】/【Select Analysis】选择需要可视化的分析项目，选择想要可视化的特征，输入相关的信息，点击【生成新图表】/【Generate New Plot】生成可视化图，如 49、图 50 所示。

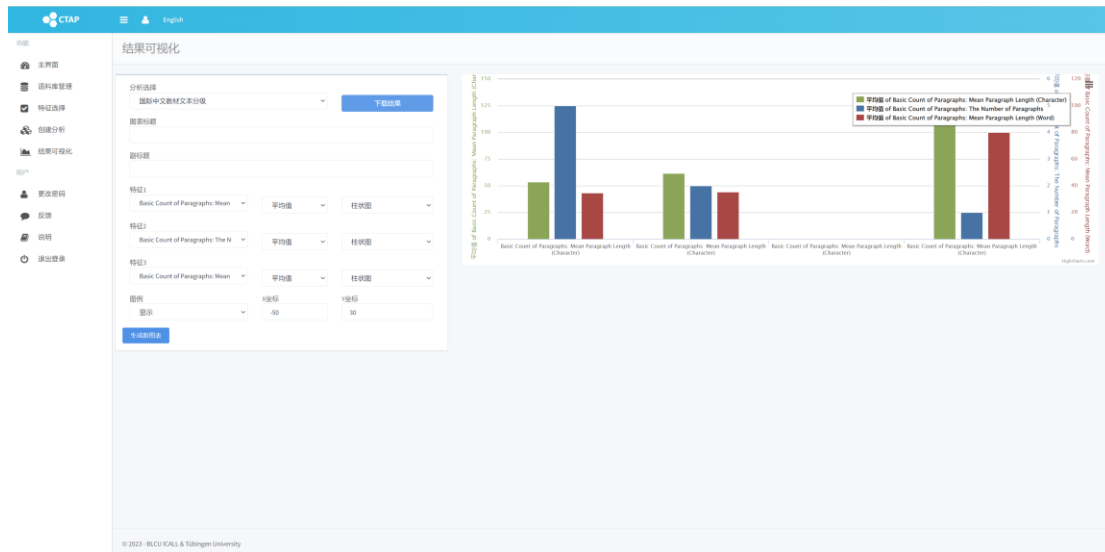


图 49 结果可视化样例-中文版



图 50 结果可视化样例-英文版