

智能化多语种教学 与科研平台 用户使用手册



目录

1. 平台简介	1
2. 语料库模块	1
2.1 教学子模块	3
2.1.1 语篇展示	3
2.1.1.1 选择语对	5
2.1.1.2 设置目录范围	5
2.1.1.3 显示/不显示译文	6
2.1.1.4 显示/不显示词性标注	6
2.1.1.5 查看词性标注对应的具体词性	7
2.1.1.6 切换显示方向	7
2.1.2 翻译策略与技巧	8
2.1.2.1 选择语对与设置目录范围	9
2.1.2.2 选择翻译策略与技巧	10
2.2 科研子模块	10
2.2.1 基础信息	11
2.2.1.1 选择统计的语对	12
2.2.1.2 选择统计的目录范围	12
2.2.1.3 查询	13
2.2.1.4 修改类符形符比基数	14
2.2.2 KWIC (Key Word In Context)	14
2.2.2.1 语对与语言	15
2.2.2.2 KWIC 目录范围	16
2.2.2.3 关键词	16
2.2.2.4 停用词	17
2.2.3 索引行	18
2.2.3.1 索引行目录范围	19
2.2.3.2 检索单个关键词	19
2.2.3.3 检索多个关键词	21
2.2.3.4 添加排除词	23
2.2.4 词频	25
2.2.4.1 设置语对与语言及目录范围	26
2.2.4.2 停用词	26
2.2.4.3 根据词性进行词频统计	27
2.2.4.4 设置统计的单词	28
2.2.4.5 跳转至 KWIC 与索引行	29
2.2.5 词簇	31
2.2.5.1 设置语对与语言及目录范围	32
2.2.5.2 关键词	32
2.2.5.3 停用词	33
2.2.5.4 词簇长度	33
2.2.6 搭配	34
2.2.6.1 设置语对与语言及目录范围	34
2.2.6.2 节点词	34
2.2.6.3 停用词	35
2.2.6.4 调整左右侧区间	35
2.2.6.5 调整检索结果	36
3. 知识库模块	38
3.1 术语库子模块	40
3.2 典故库子模块	41
4. 文献库模块	41

1. 平台简介



The banner features the Shanghai International Studies University (SISU) logo and the Corpus Studies and Applications Institute logo on the left. The text '您好, 李晓倩' is on the right. Two cards are displayed: the left one is titled '《习近平谈治国理政》多语数据库综合平台' and the right one is titled '自定义语言数据分析与应用平台'. Both cards include brief descriptions of their functions.

© 2022 上海外国语大学语料库研究院

智能化多语种教学科研平台由上海外国语大学语料库研究院胡开宝教授领衔的团队研发，该平台旨在推进语料库在外语教学与科研中的应用，实现语言教学与科研的数据化、可视化与智能化。

智能化多语种教学科研平台由《习近平谈治国理政》多语种数据库综合平台和自定义语言数据分析与应用平台组成，本平台使用说明将以《习近平谈治国理政》多语种数据库综合平台为例。

智能化多语种教学科研平台包含语料库（教学科研平台）、知识库、文献库三大主要功能模块。

2. 语料库模块

平台首页单击语料库模块，即进入相应语料库模块页面。



图 1



图 2

目前，语料库模块内的语料数据包括《习近平谈治国理政》第一至第三卷中文原文及其对应英语译文。

针对外语学科教学与科研工作的具体需求，平台在‘语料库’模块下设有教学子模块与科研子模块，但在具体使用时，用户不必局限于单个子模块的功能，应当结合个人或团队应用场景的具体需求，有机融合两个子模块的功能。在本模块使用说明的最后，将会介绍两个教学与科研子模块搭配使用的应用案例。

2.1 教学子模块

在语料库模块首页点击教学模块。



图 3

教学子模块的主要功能为：1.语篇展示；2.翻译策略与技巧。



图 4

2.1.1 语篇展示

在教学子模块首页点击语篇展示即可进入语篇展示功能页。



图 5

语篇展示功能可以显示单语语篇或句级对齐的双语或多语平行语篇，以方便在外语教学过程中了解某一特定语篇的篇章内容，用户可按照自身需求来展示有或没有译文对照的《习近平谈治国理政》不同语言版本的语篇内容，与纸质版相比，语篇展示功能能够让用户更加快速、更加便捷地查看相应的章节。其中，右上角的“切换目录”功能，可以满足对《习近平谈治国理政》不同章节内容的选择与切换。

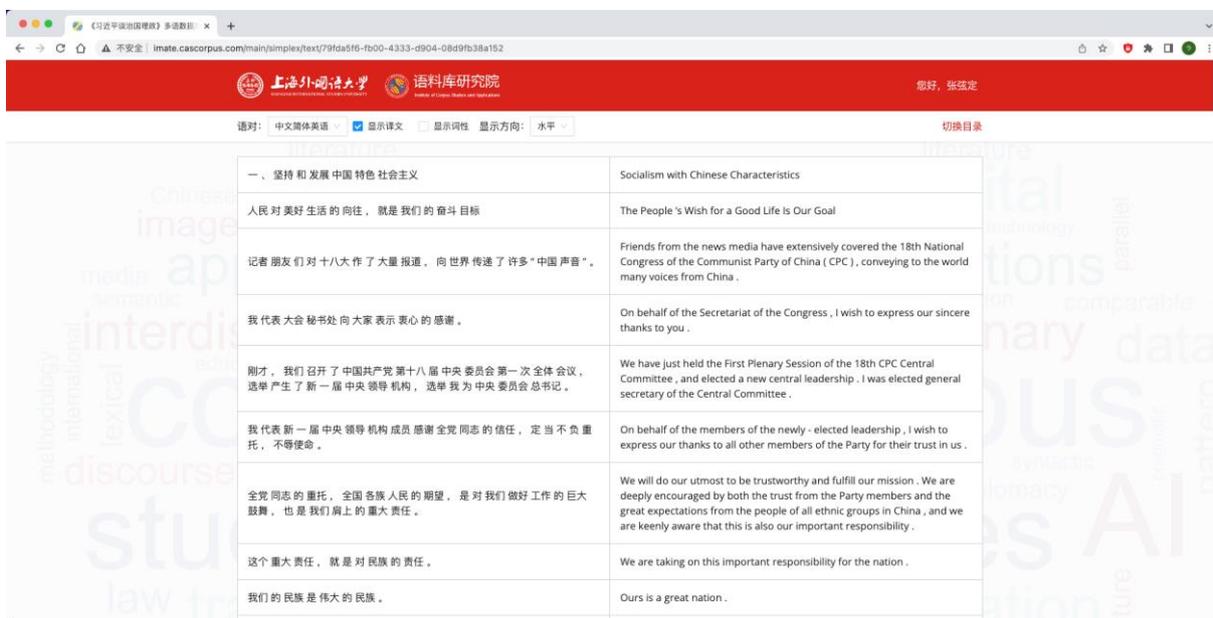


图 6

2.1.1.1 选择语对

在使用语篇展示功能时，用户可以按照需求选择相应的语对，平台会自动展示对应语种的语篇内容，如图 7 所示。

目前，本地部署后的平台 2.0 版本最多可以支持多个语对的语篇展示。

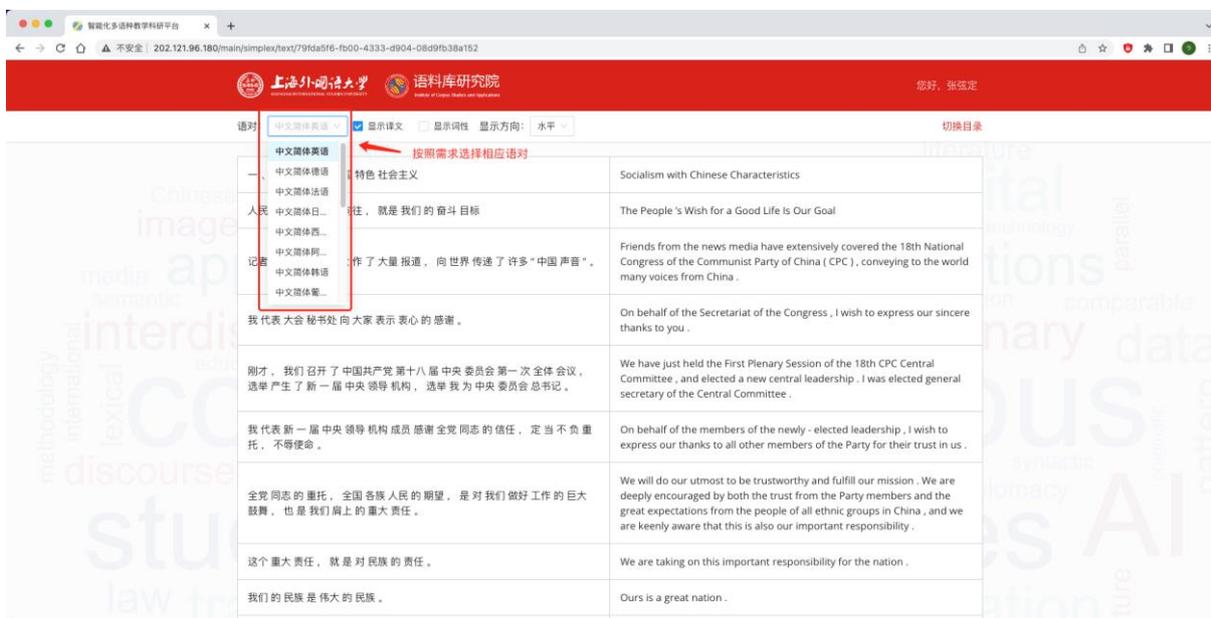


图 7

2.1.1.2 设置目录范围

点击右上角切换目录，平台会显示《习近平谈治国理政》多语种版本前 3 卷的目录，用户可以按需选择。

以图 8 所示为例，《习近平谈治国理政》多语种版本目录可视为一个三级目录，第一卷可视为一级标题，坚持和发展中国特色社会主义可视为二级标题，人民对美好生活的向往，就是我们的奋斗...可视为三级标题，选择一级标题则会展示该一级标题下所有二级、三级标题内的语篇，以此类推。

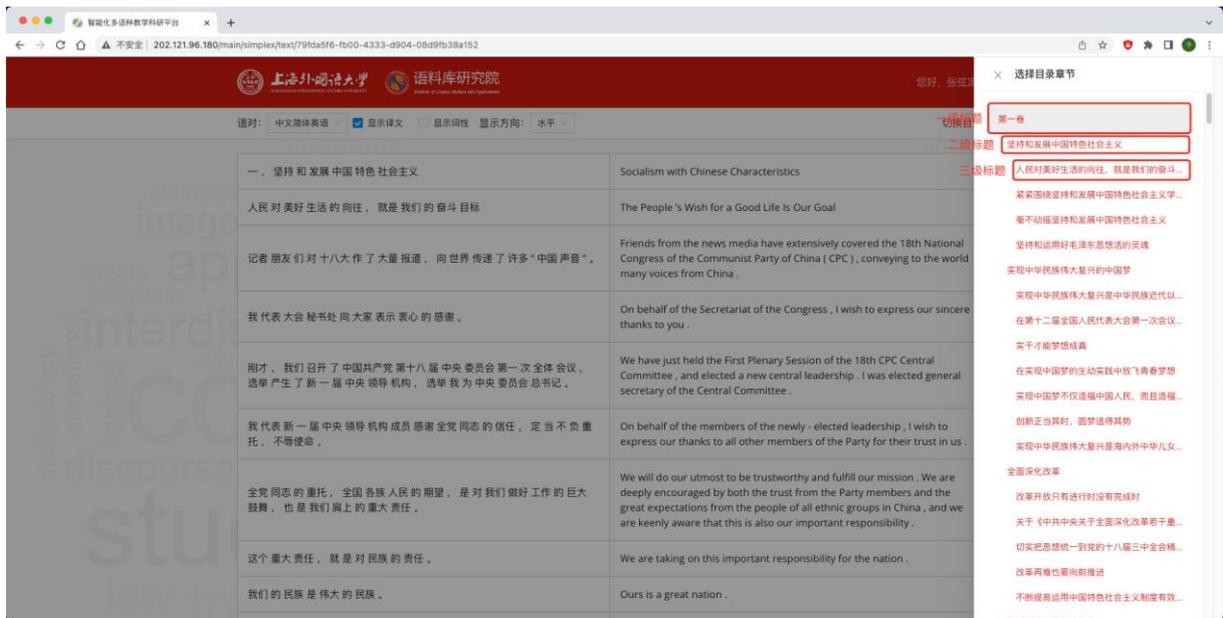


图 8

在使用平台的大多数功能时，都需要先根据自身的教学与科研需求设置好目录范围，在后续的说明中将不再对设置目录范围进行赘述，操作方法可参照本节。

2.1.1.3 显示/不显示译文

取消勾选显示译文，平台会自动隐藏译文的语篇内容，如图 9 所示。

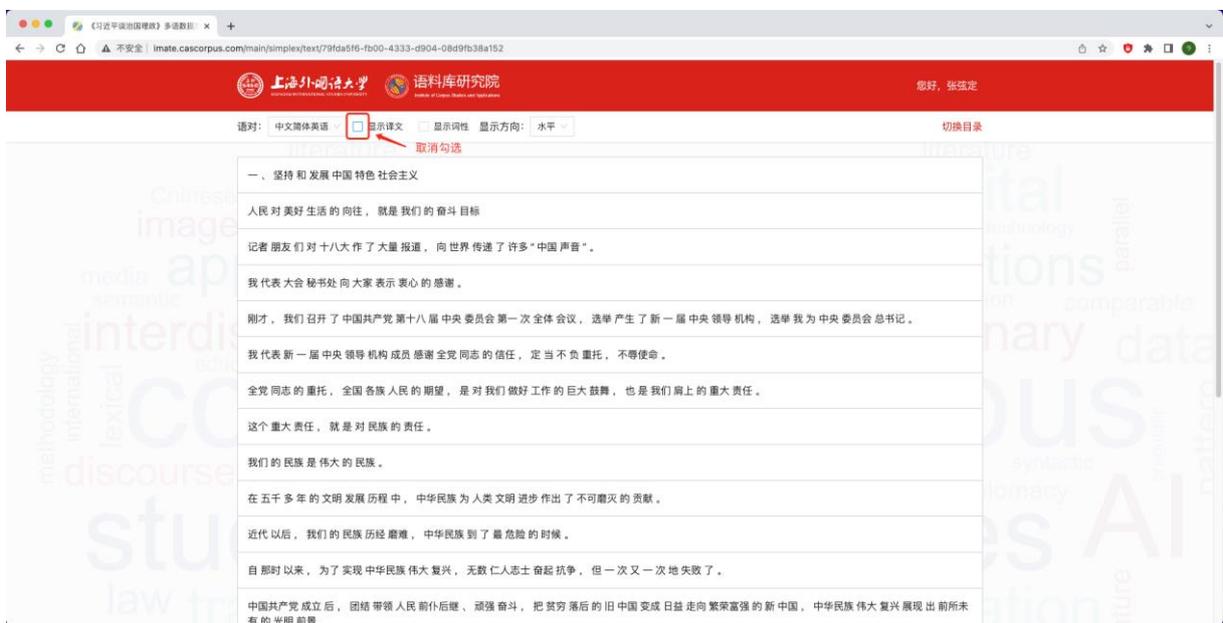


图 9

2.1.1.4 显示/不显示词性标注

勾选显示词性，平台会显示每一个词对应的词性标注，如图 10 所示。



图 10

2.1.1.5 查看词性标注对应的具体词性

通过将鼠标移至词性标注上，可以查看词性标注对应的具体词性，如图 11 所示。

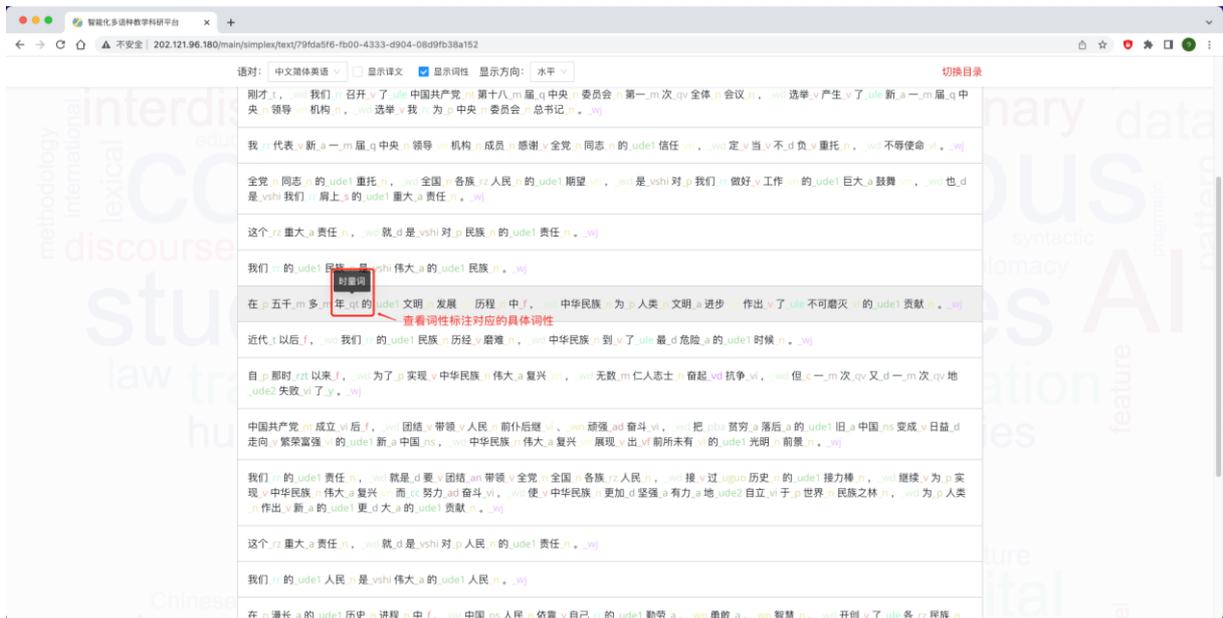


图 11

2.1.1.6 切换显示方向

在显示译文的状态下，用户可点击‘显示方向’下拉框，将显示方向切换为垂直，如图 12 所示。

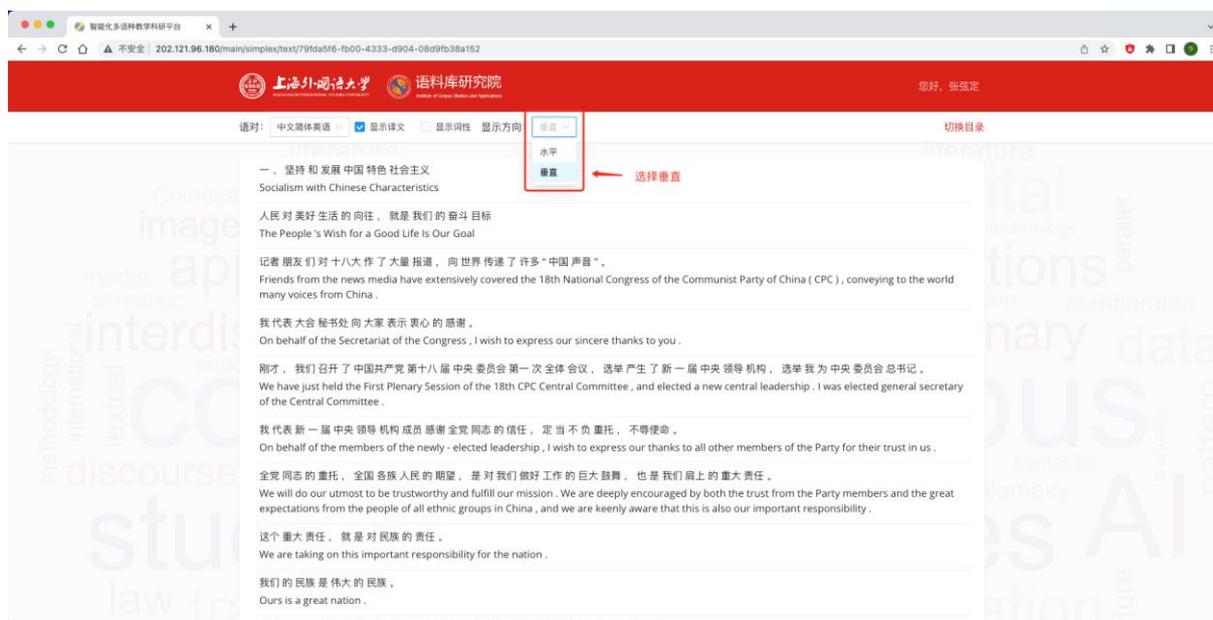


图 12

2.1.2 翻译策略与技巧

在教学子模块首页点击翻译策略与技巧即可进入翻译策略与技巧功能页。



图 13

翻译策略与技巧查询模块可根据不同目录、语对以及翻译策略与技巧等进行相应查询，考察特定语境中具体词汇、短语等表述的翻译策略与技巧，为翻译教学与科研提供丰富的翻译实例。

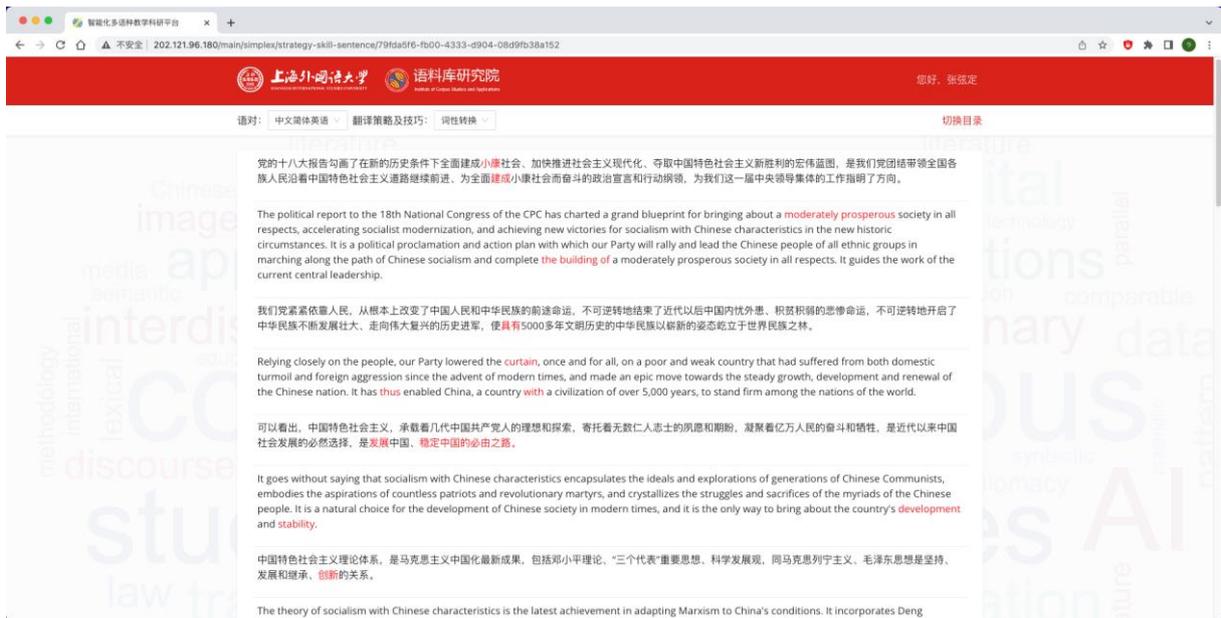


图 14

2.1.2.1 选择语对与设置目录范围

翻译策略和技巧和语篇展示一样，同样支持选择语对与切换目录，具体操作方法可参照 [2.1.1.1 选择语对](#)和 [2.1.1.2 设置目录范围](#)。



图 15

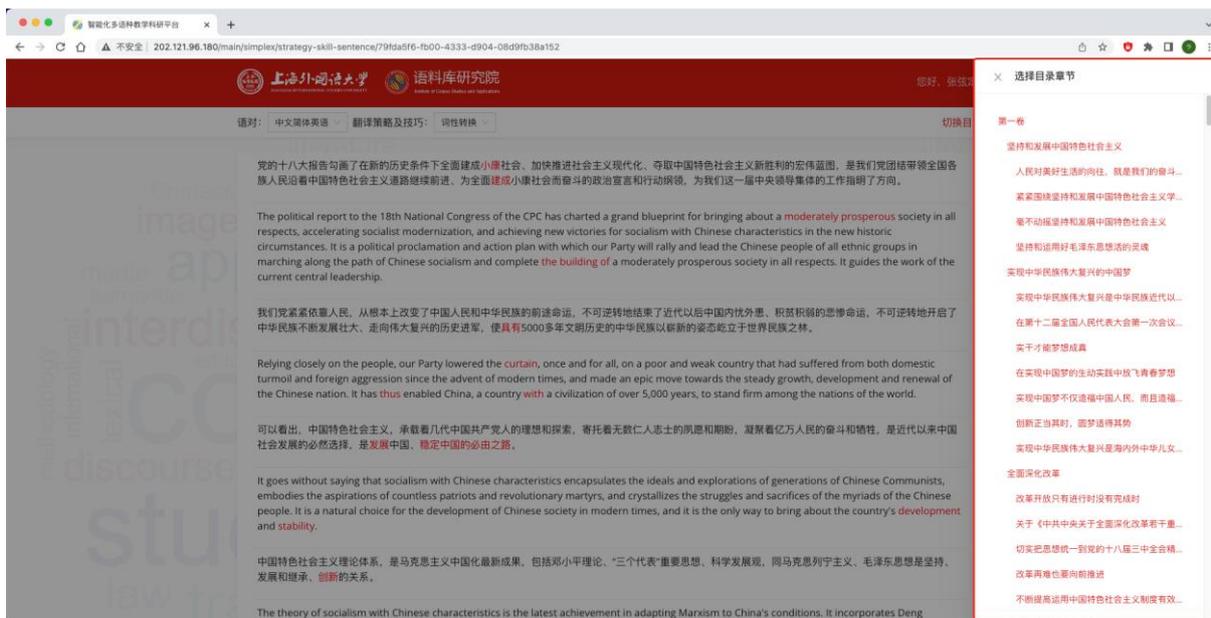


图 16

2.1.2.2 选择翻译策略与技巧

点击翻译策略与技巧下拉框，选择需要查看的翻译策略与技巧，平台会展示采用了对应翻译策略或技巧的中外平行语篇，如图 X 所示，语对为“中文简体英语”，目录为“第一卷”，翻译策略与技巧选择“视角转换”，平台会展示第一卷中所有采用了“视角转换”的中英平行语篇。

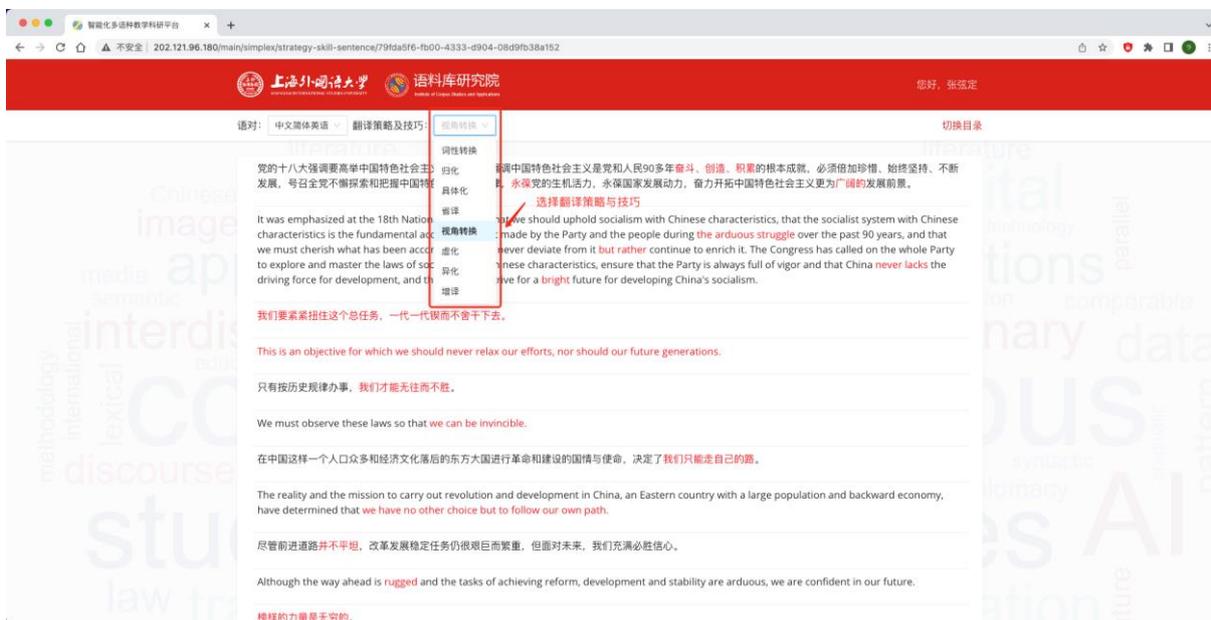


图 17

2.2 科研子模块

在语料库模块首页点击科研模块。



图 18

利用科研子模块，用户可以进行单语或平行语料的检索，同时，科研子模块还提供基础信息、KWIC、索引行、词频、词簇、搭配等语言数据的查询，并在提供具体检索结果的基础上，实现词频、搭配等数据结果的可视化呈现。

2.2.1 基础信息

基础信息功能提供所选语料的类符、形符、类符形符比、标准化类符形符比等总体语言数据。

类符 (type) 是指语料库中的不同词汇，或每个第一次单独出现的词形。

形符 (token) 是指语料库中出现的所有词形。

类符/形符比 (TTR, type/token ratio) 指文本中类符与形符的比率，计算公式为：

$$TTR = \text{type} / \text{token} * 100\%$$

标准化类符/形符比 (STTR, standardized type/token ratio) 的计算方法为：按照一定长度（如 1000 形符）切分文本，再逐个计算各切分文本的类符/形符比，最后取其平均值，即为标准化类符/形符比。

设定中文简体/英语语对，目录范围为一至三卷全部，切分文本的类符形符比基数为 1000 形符，统计结果如图 19 所示。

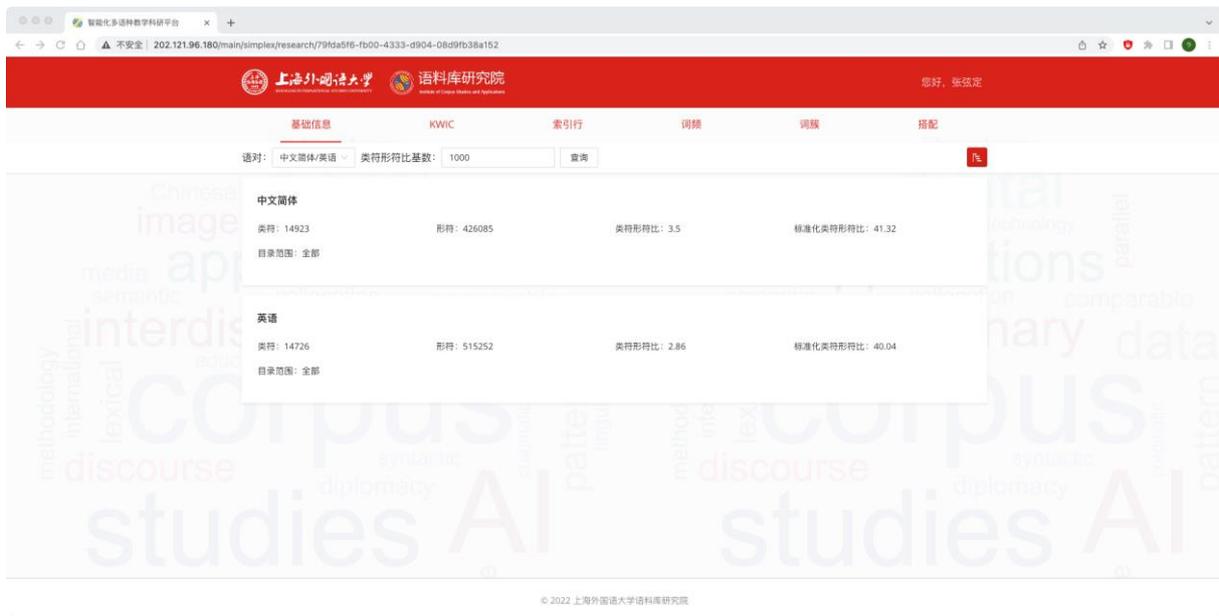


图 19

2.2.1.1 选择统计的语对

点击语对下拉框，设置需要统计的语对，此处以中文简体/德语为例。



图 20

2.2.1.2 选择统计的目录范围

点击右上角红色图标，设置需要统计的篇章范围，具体操作方法可参照 [2.1.1.2 设置目录范围](#)。默认范围为全部篇章，此处以选择第一卷—坚持和发展中国特色社会主义为例。

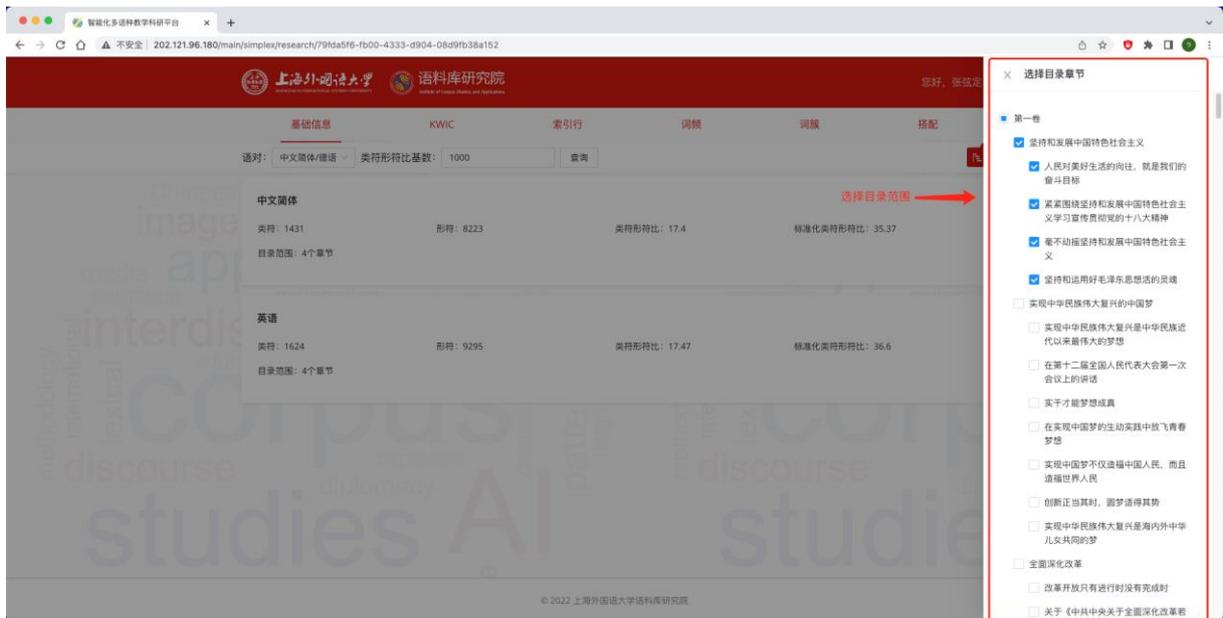


图 21

2.2.1.3 查询

设置好语对与目录范围后，点击查询键，平台会依照条件进行统计并显示结果。

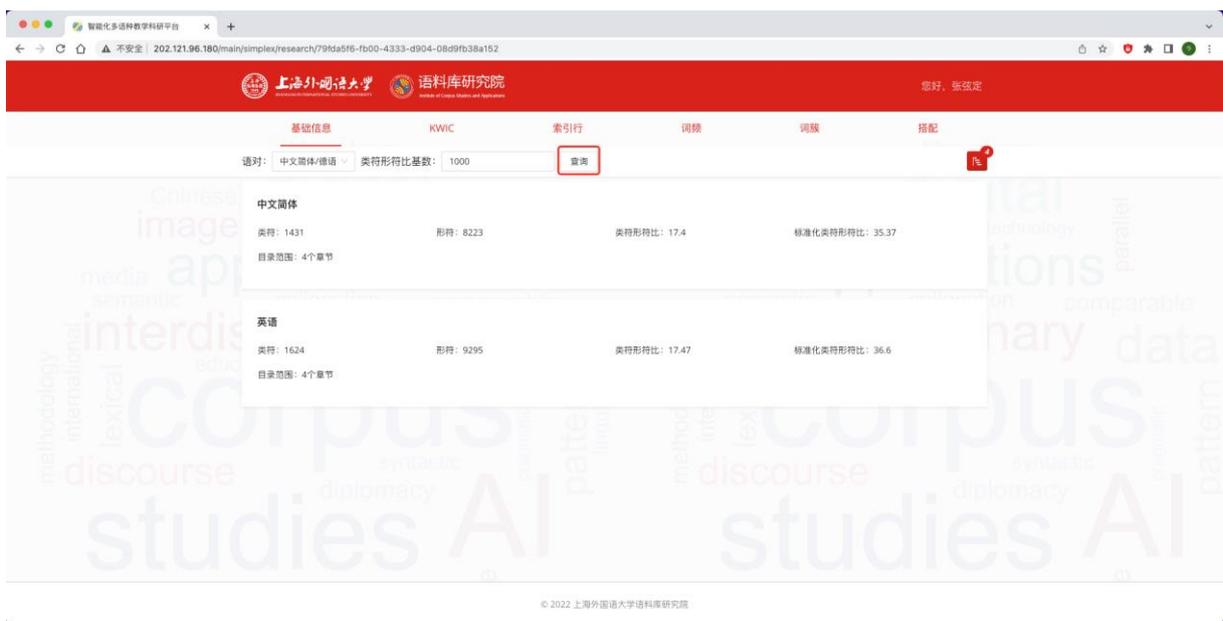


图 22

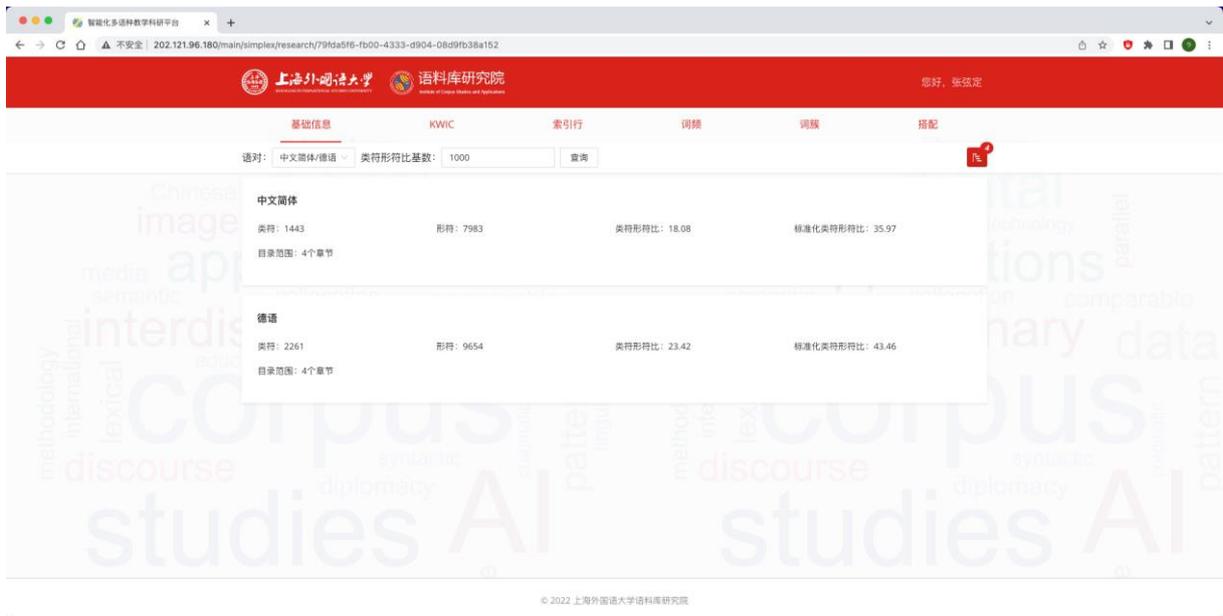


图 23

2.2.1.4 修改类符形符比基数

用户可以根据所选目录范围内的形符类符数，结合自身的科研需求设定类符形符比基数，平台会自动根据基数切分语料并计算标准化类符形符比。

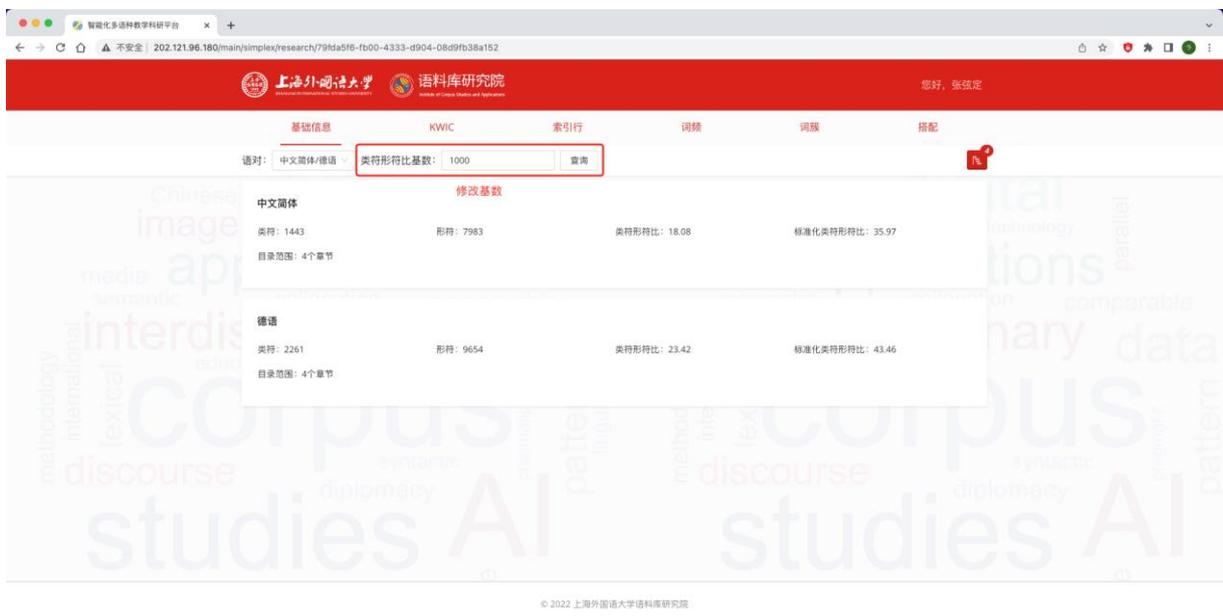


图 24

2.2.2 KWIC (Key Word In Context)

利用 KWIC 功能，用户可以查询具体词汇的上下文使用语境。

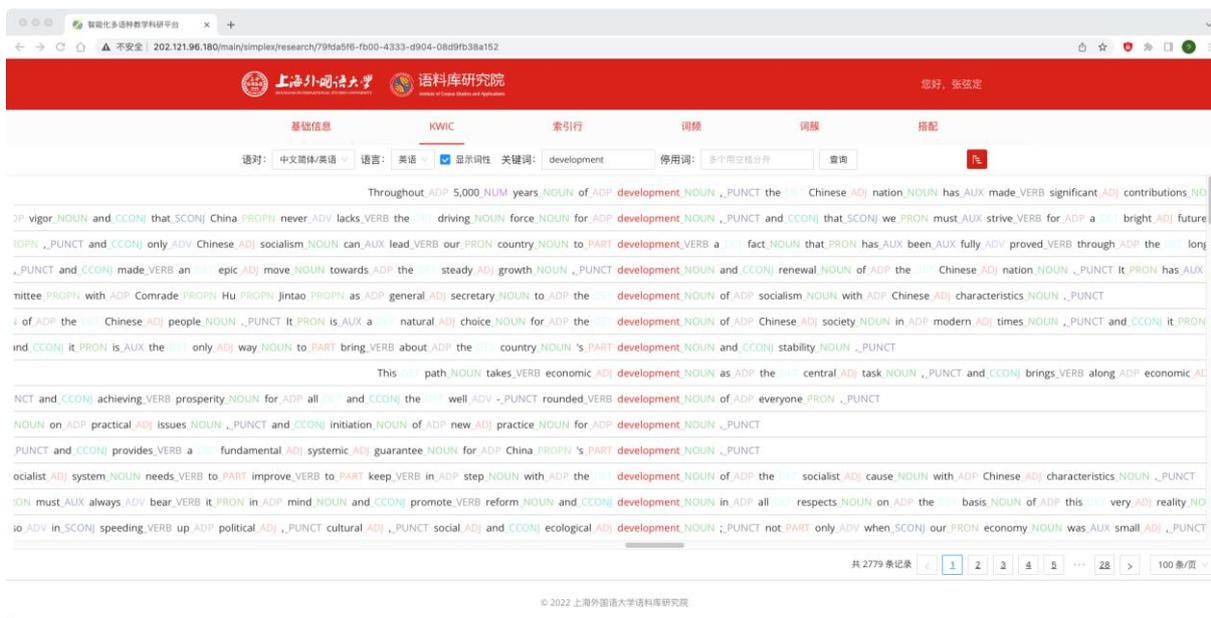


图 27

2.2.2.2 KWIC 目录范围

KWIC 功能同样支持设置目录范围，具体操作请参照 [2.1.1.2 设置目录范围](#)，基于上述检索结果，点击右上角红色图标，选择坚持和发展中国特色社会主义，则可以限定在该章节范围内进行检索。

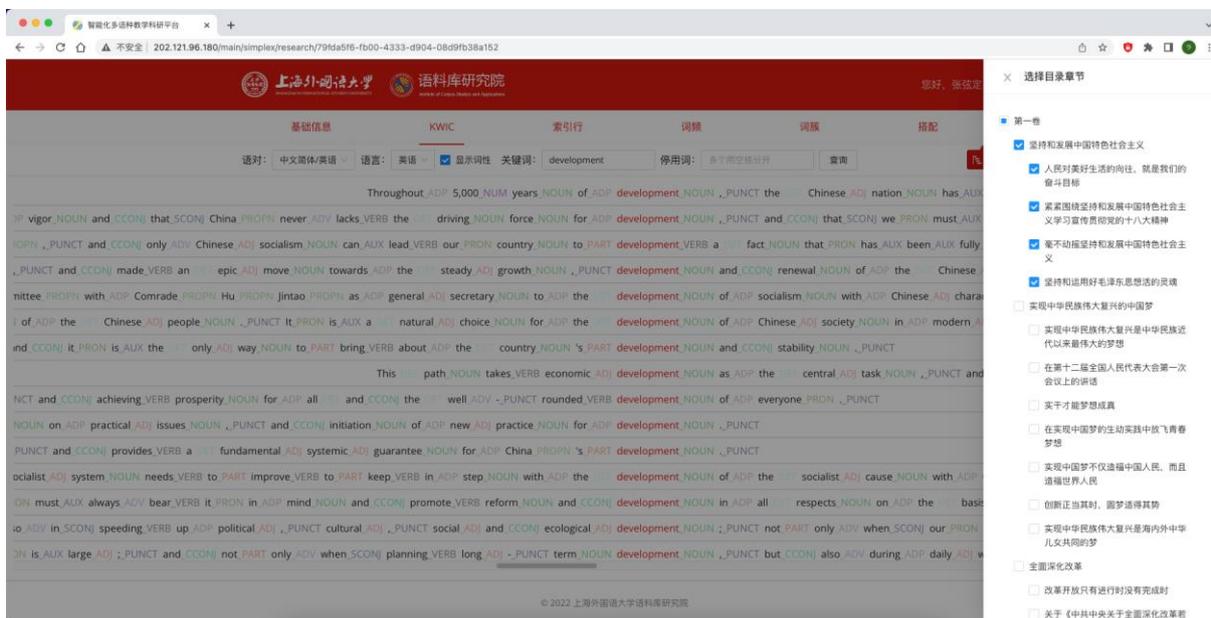


图 28

2.2.2.3 关键词

使用关键词前首先需要确定关键词的语言，具体操作范围请参照 [2.2.2.1 语对与语言进行设置](#)。

以单词 economic 为例，语对应设置为中文简体/英语，语言应设置为英语，随后在关键词栏输入 economic，目录范围及词性显示根据具体需求设定，此处以第一卷-坚持和发展中国特色社会主义，显示词性为例，点击查询即可显示检索结果，如下图所示。

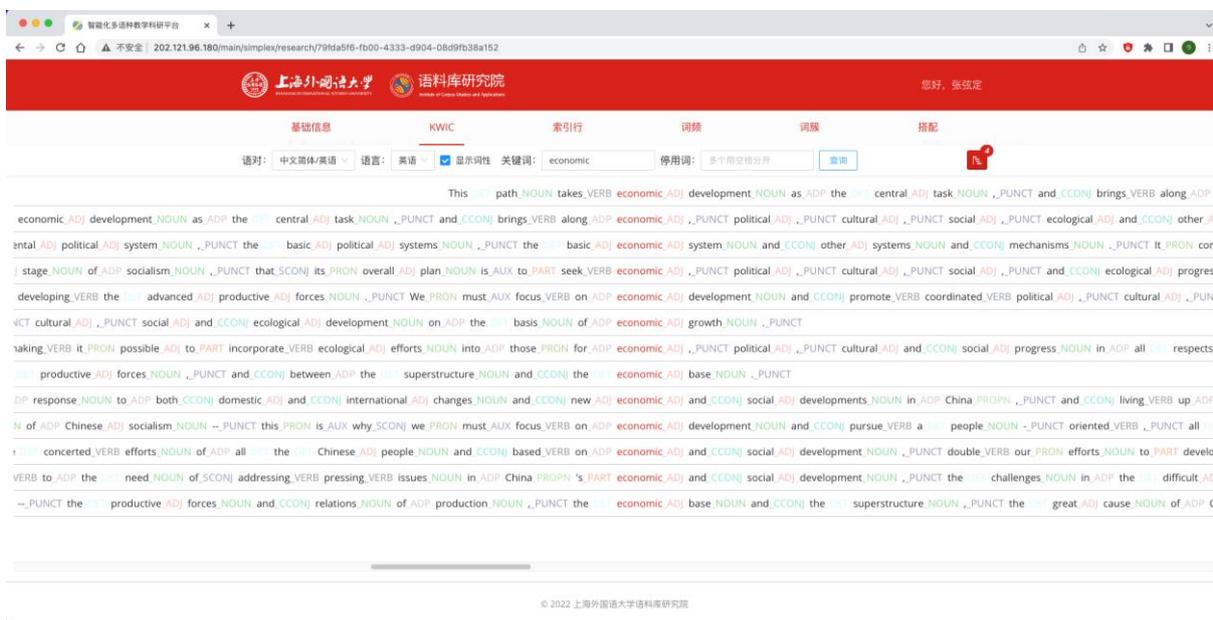


图 29

关键词部分还支持使用通配符进行检索，用户可根据需要自行构建通配符，例如，如需检索所有以 en-开头的单词，只需在关键词栏中输入 en*，然后点击查询：

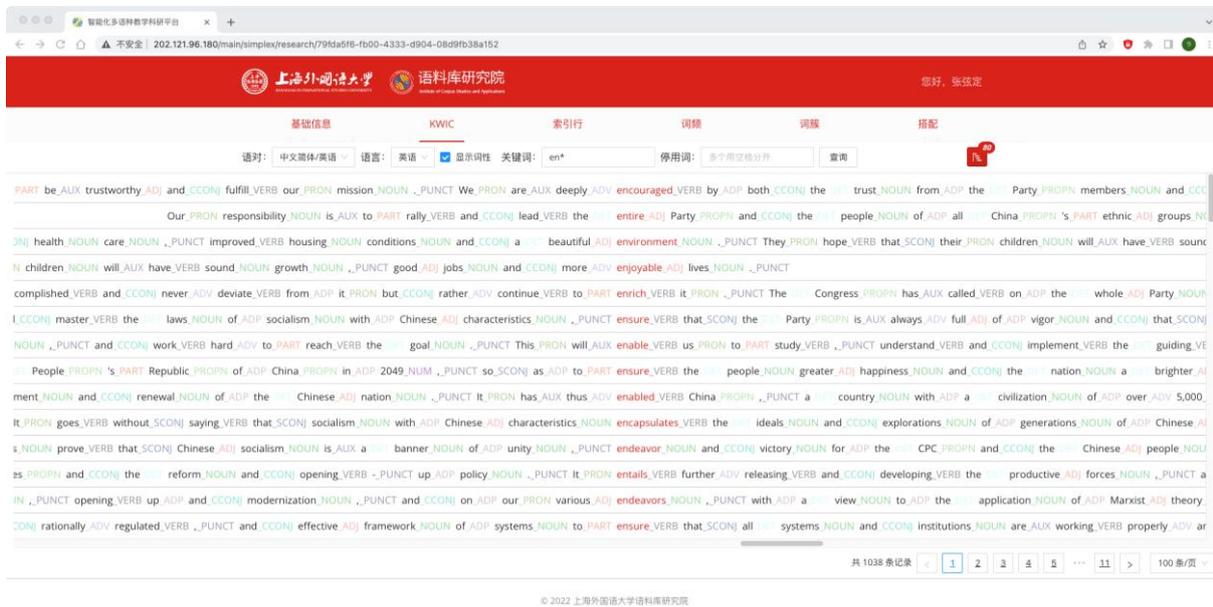


图 30

2.2.2.4 停用词

基于 2.2.2.3 关键词的检索结果，用户还可以根据需求设置停用词，平台会在检索结果中剔除

含有停用词的结果。

此处以停用 as 为例，用户只需在停用词栏中输入 as，然后点击查询：

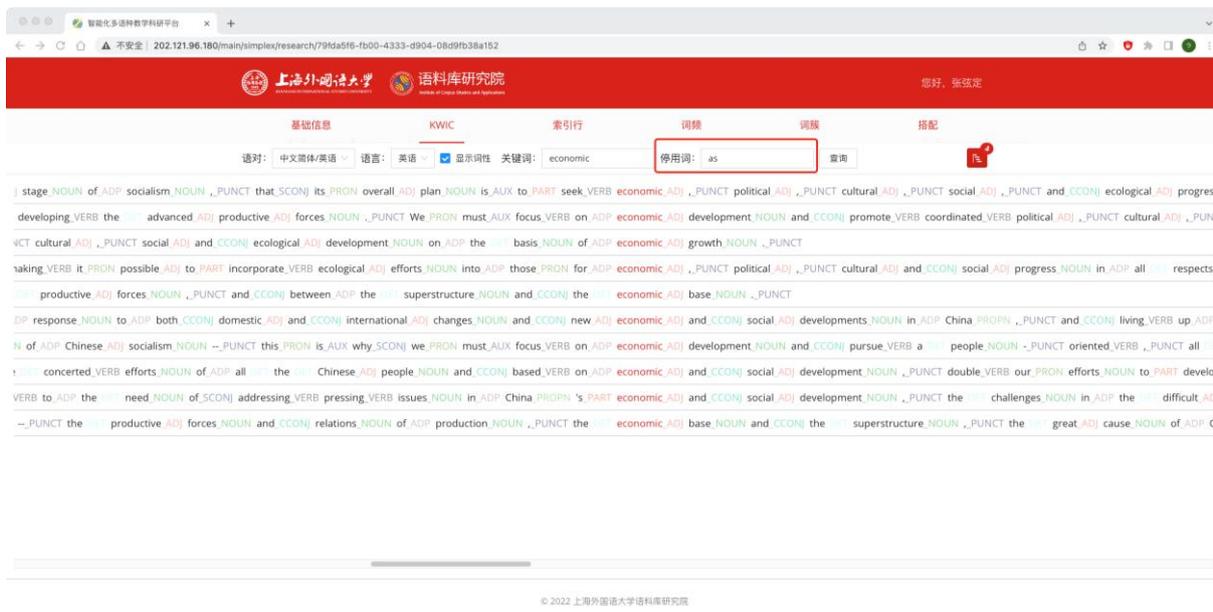


图 31

平台也支持同时设置多个停用词，多个停用词之间需用空格隔开，以停用 as 和 to 为例，在停用词栏中输入 as to，然后点击查询：



图 32

2.2.3 索引行

索引（concordance）指利用搜索功能，将某词语及其在文本中的语境一同逐行呈现出来。使用索引功能时，被检索的词称为检索词，索引结果以句子形式呈现。可以说，索引行展示了检索词的实际应用情况，尤其提供了检索词所在的语境。

平台的索引行模块可以按提供的检索词，进行单语或平行索引行的检索，展现符合特定检索条

件的句子，检索结果可以以单语或多语形式显示。该模块支持复合检索和排除检索，如检索同时包含“环境”和“经济”的句子。索引行模块可以为语言教学与科研提供大量、真实的语言使用实例，为不同语种语言的观察、分析等提供语料基础。

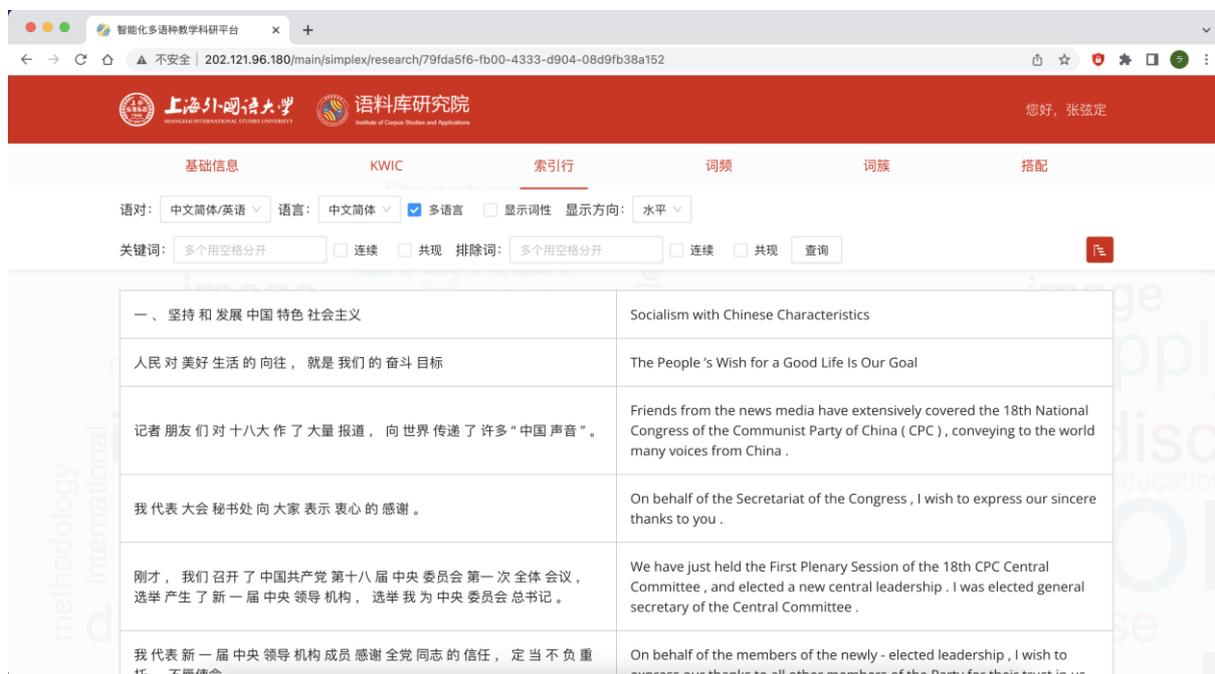


图 33

2.2.3.1 索引行目录范围

在使用索引行功能前，用户需根据教学与科研需求设定检索的目录范围，具体操作方法请参照 [2.1.1.2 设置目录范围](#)。

2.2.3.2 检索单个关键词

在关键词栏输入关键词，关键词的语种设置可参照 [2.2.2.1 语对与语言](#)，点击查询即可在选定的目录范围内进行检索，平台默认显示双语平行的检索结果，以检索“环境”，目录范围为全部为例，结果如下：

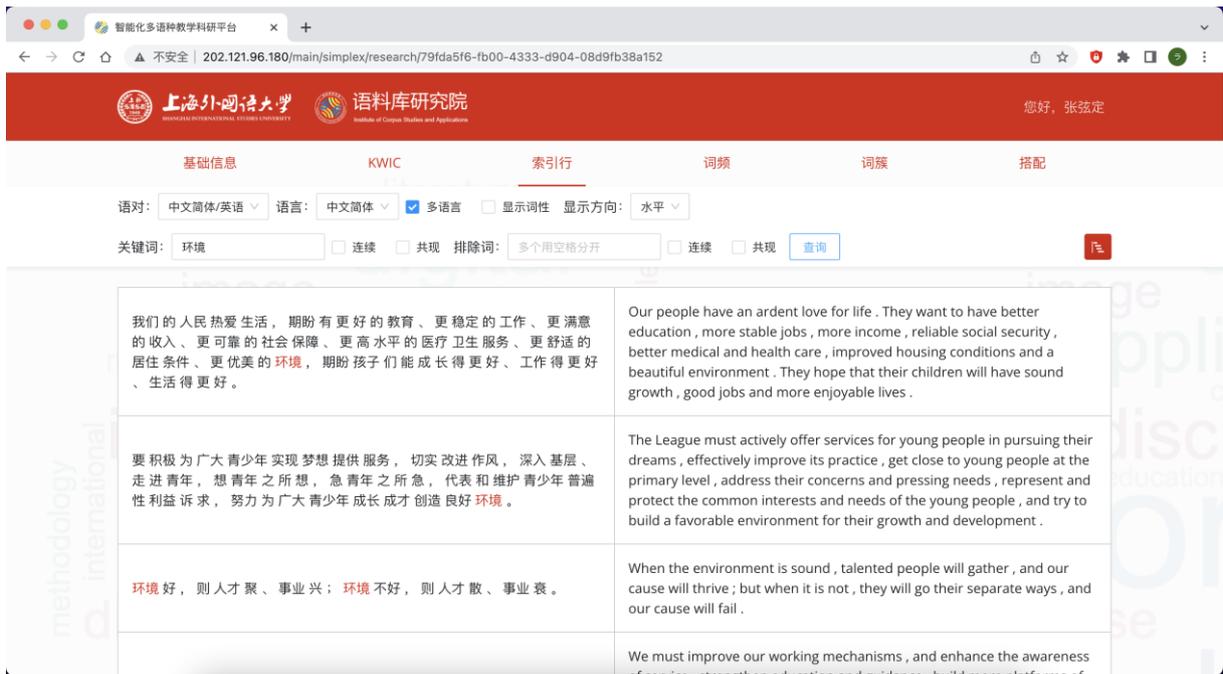


图 35

根据需要，可取消勾选多语言，平台将显示单语检索结果：

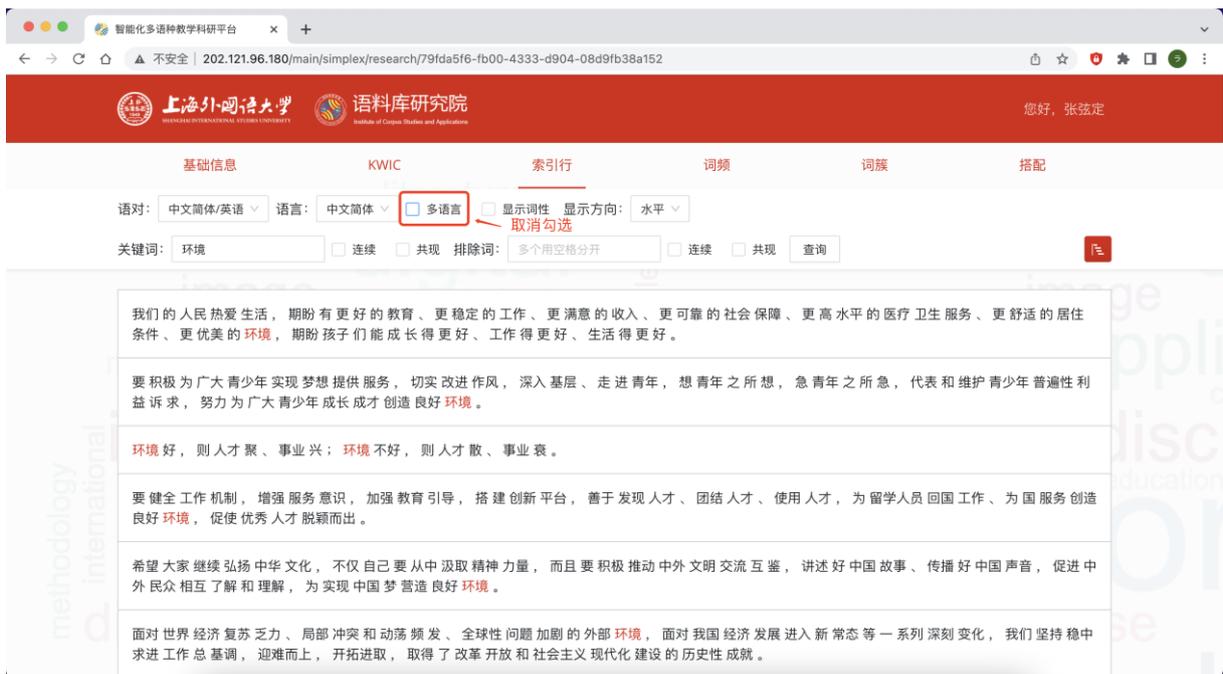


图 36

检索行功能支持在检索结果中显示词性：

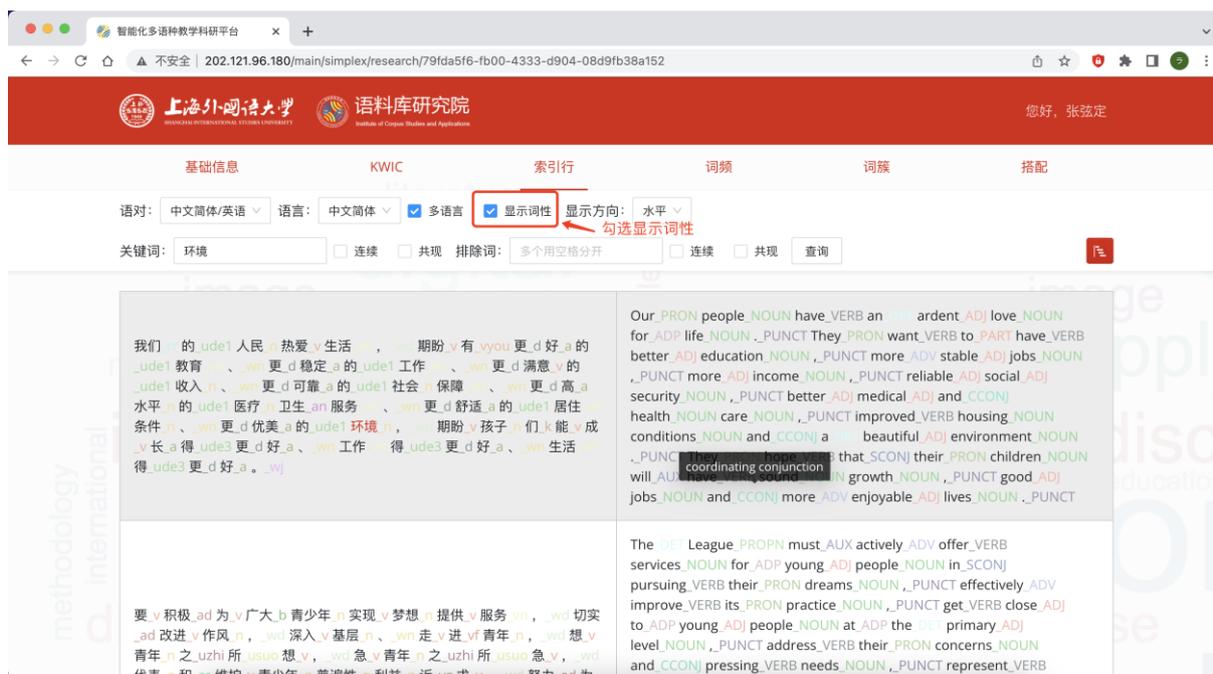


图 37

检索结果可以垂直显示：

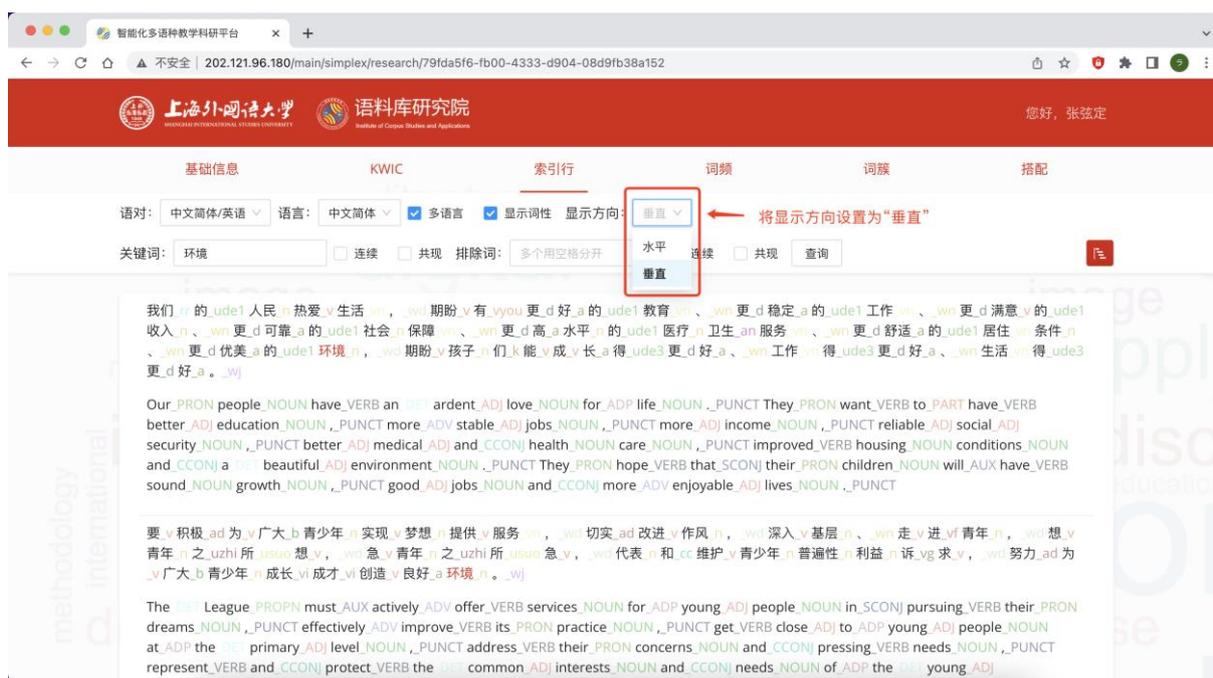


图 38

2.2.3.3 检索多个关键词

用户可在关键词栏输入多个关键词，用空格隔开，以对多个关键词进行检索，以“环境 发展”为例，平台会检索含有“环境”以及含有“发展”的内容，检索结果如下：

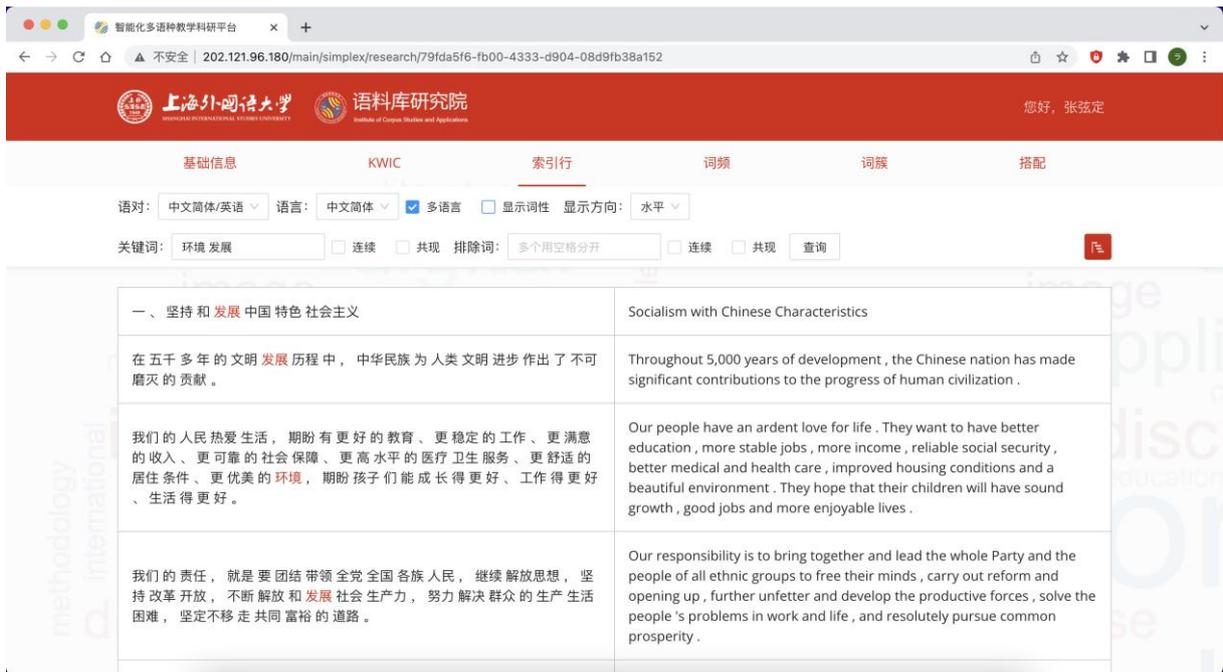


图 39

基于上述检索结果，如果用户需要查看包含“环境发展”的内容，可以勾选“连续”，平台会将“环境”与“发展”视为一个整体进行检索，检索结果如下：

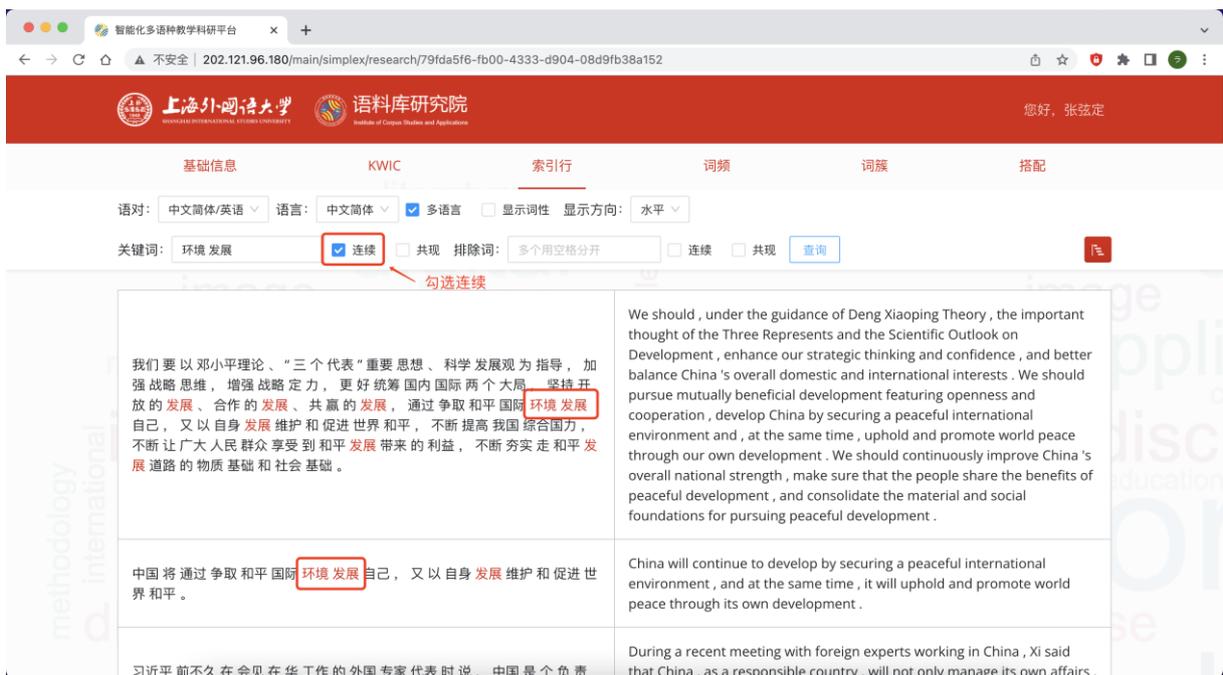


图 40

如果需要查看同时包含“环境”与“发展”的内容，用户可以勾选“共现”，检索结果如下：

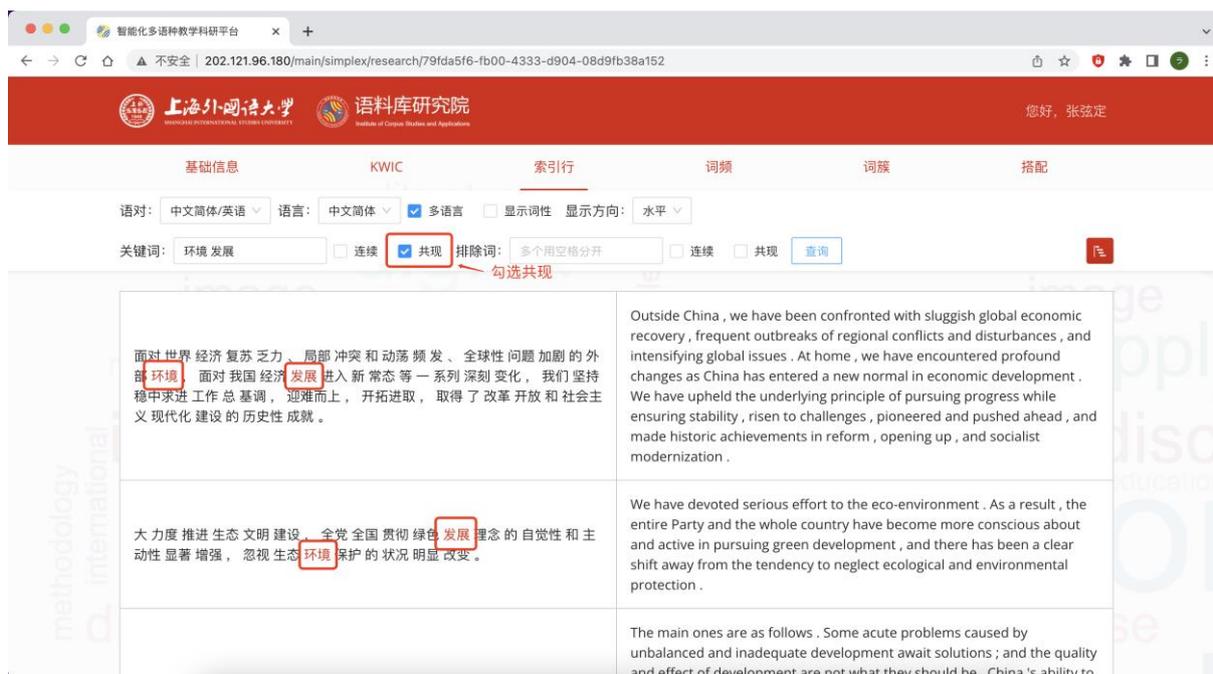


图 41

平台也支持同时满足“连续”与“共现”条件的检索：

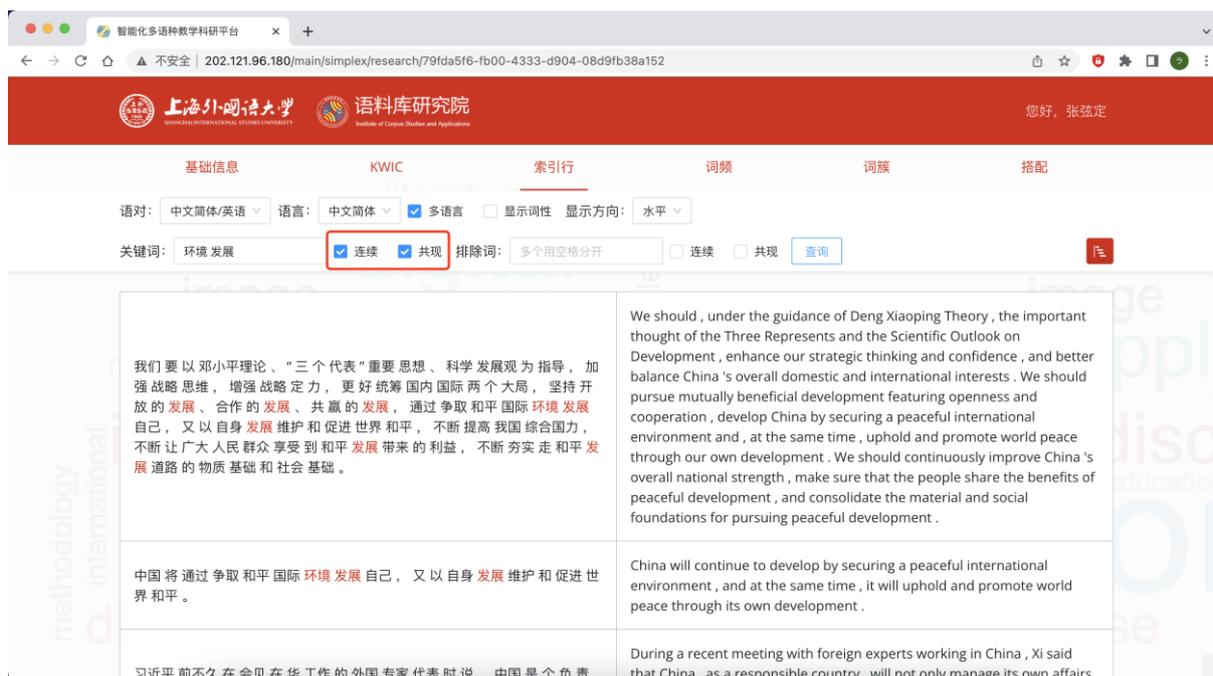


图 42

2.2.3.4 添加排除词

基于 2.2.3.2 中图 42 的检索结果，用户添加排除词，以排除包含对应词汇的内容，以排除“社会”为例：

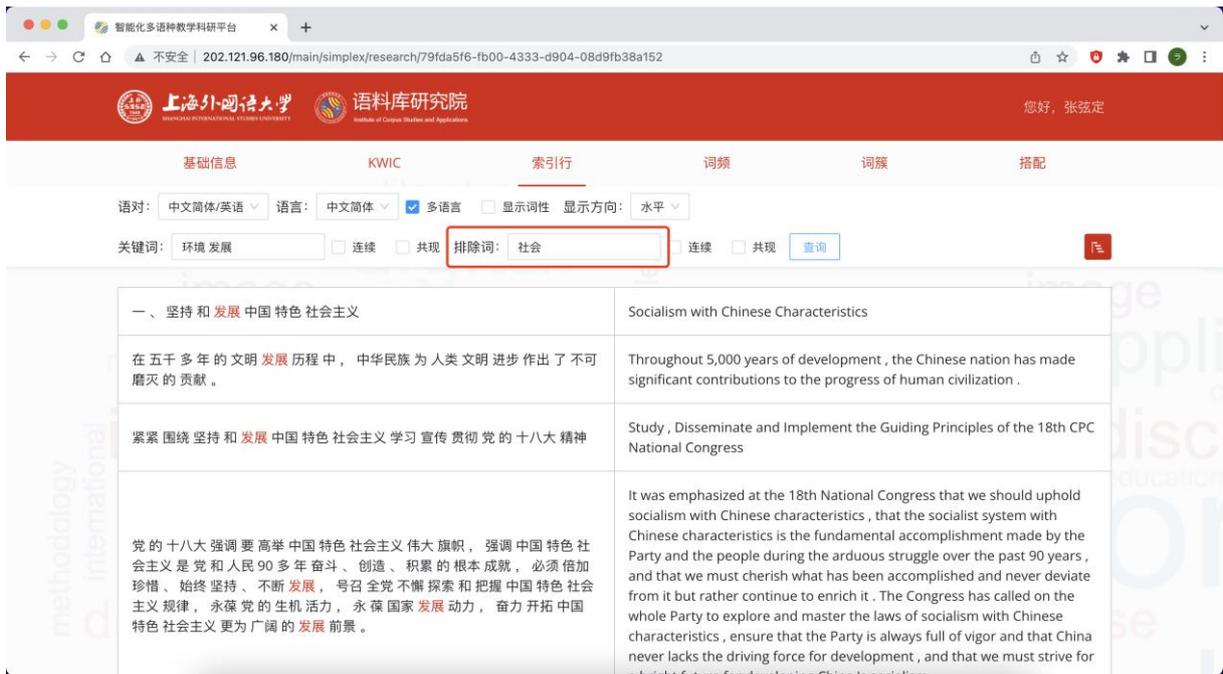


图 43

备注：在上述检索结果中，有些语句含有“社会”但仍出现在结果中，原因是在这些内容中，“社会”并不单独成词，而是包含在“社会主义”这一个整体中，因此索引行的检索并不会将其排除在外。

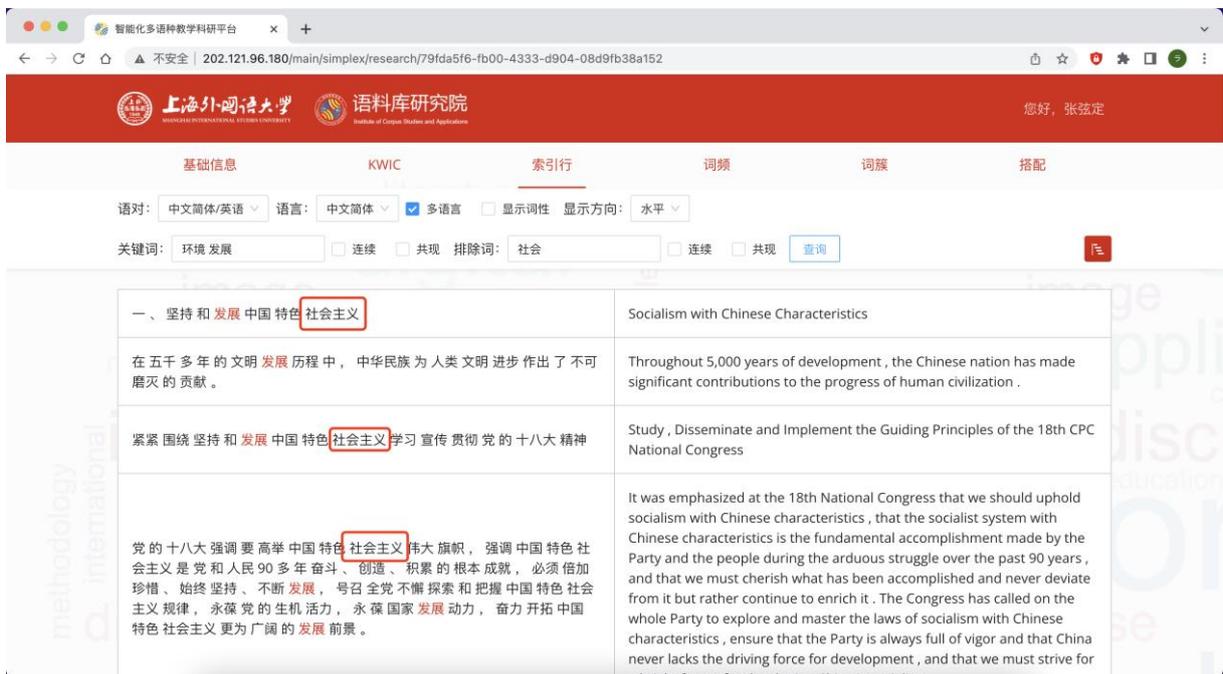


图 44

排除词也支持多个关键词，以“社会”与“进步”为例：

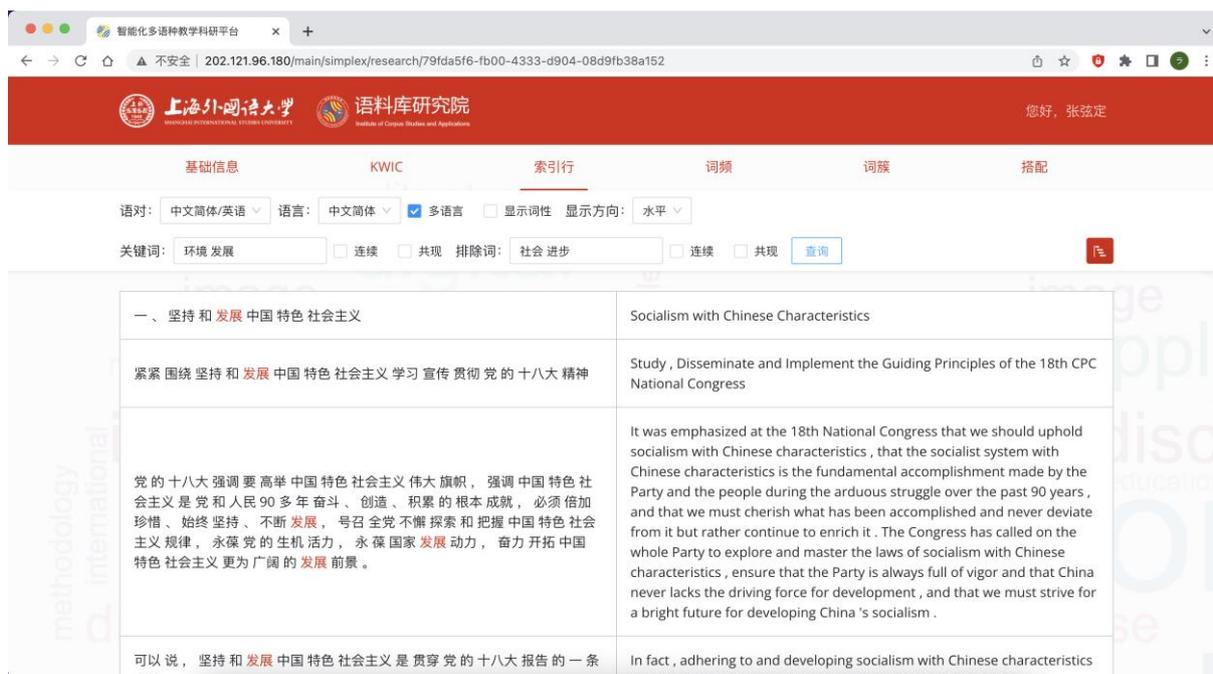
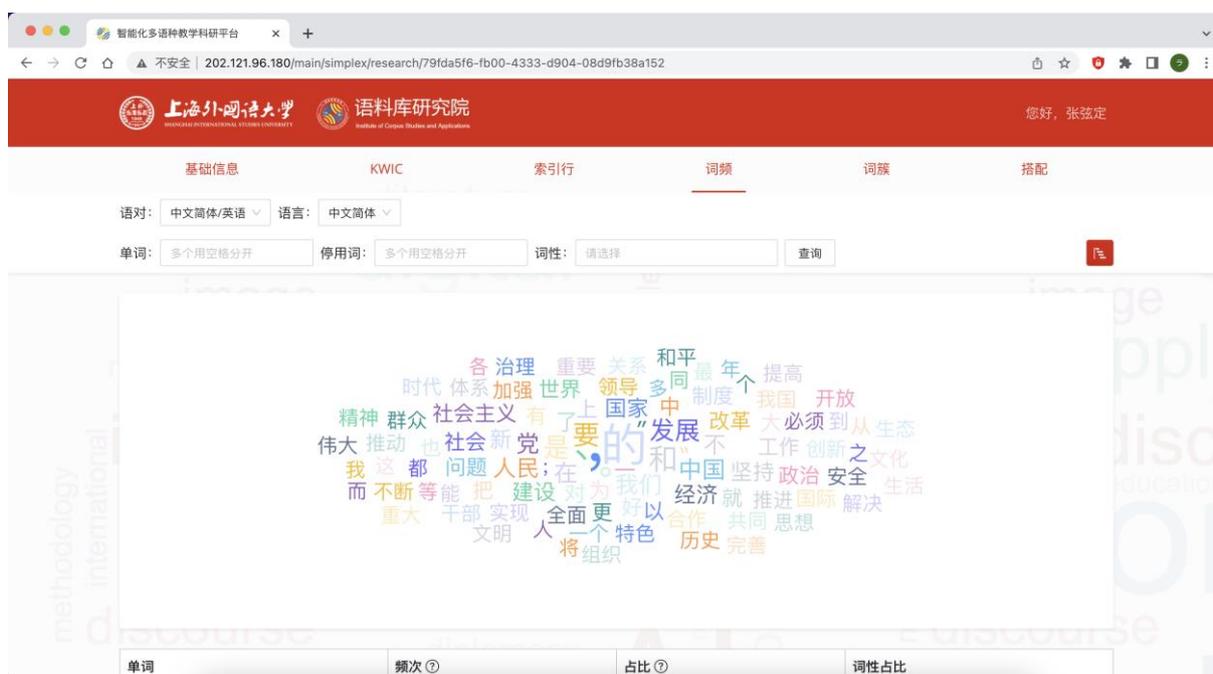


图 45

排除词同样支持“连续”与“共现”的检索，具体操作方法请参照 2.2.3.2 检索多个关键词中有关“连续”与“共现”的说明。

2.2.4 词频

词频模块可以展示选定语篇中词汇的出现频次、占比、标准化频率等信息，并对其进行可视化呈现。平台支持选择某一类或多类词性、排除某些词汇或排除特定词性的词等多个设置。词频的查询结果以可视化云图以及具体表格两种形式展示。



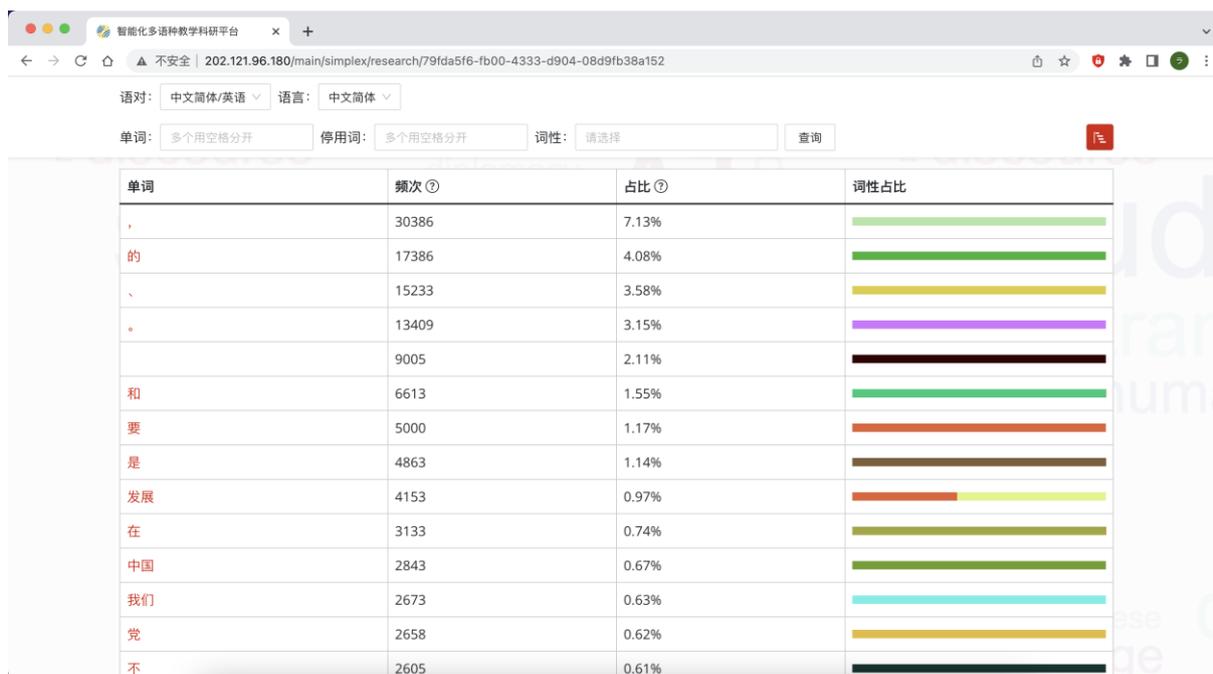


图 47

2.2.4.1 设置语对与语言及目录范围

具体操作请参照 [2.2.2.1 语对与语言](#) 及 [2.1.1.2 设置目录范围](#)。

2.2.4.2 停用词

在默认情况下，词频统计中占据前几位的一般为虚词及标点符号，如需避免这种情况，使得词频统计结果更加适用于自身的教学与科研工作，用户可以使用停用词功能。

在停用词栏中输入需要停用的单词或标点符号，中间用空格隔开，平台将不会对停用词进行词频统计，以停用“，”、“的”、“、”、“。”、“和”为例，词频统计结果如下：

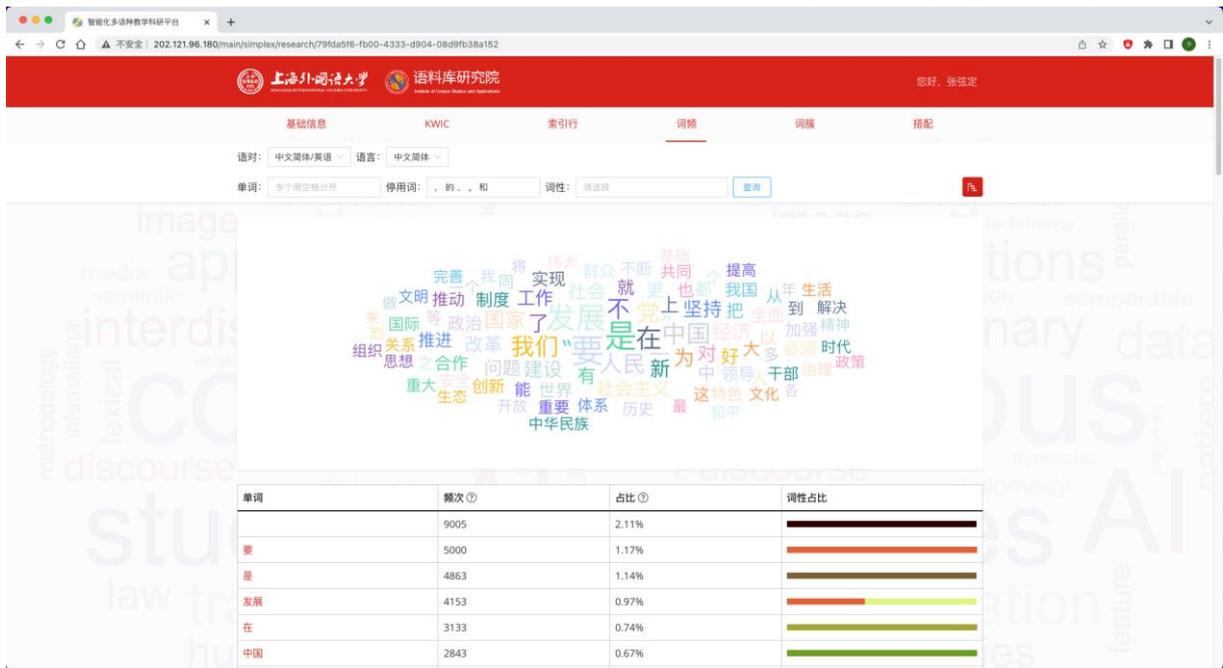


图 48

2.2.4.3 根据词性进行词频统计

如需按照词性进行词频统计，可以在词性栏中进行筛选，首先点击词性栏的下拉框。

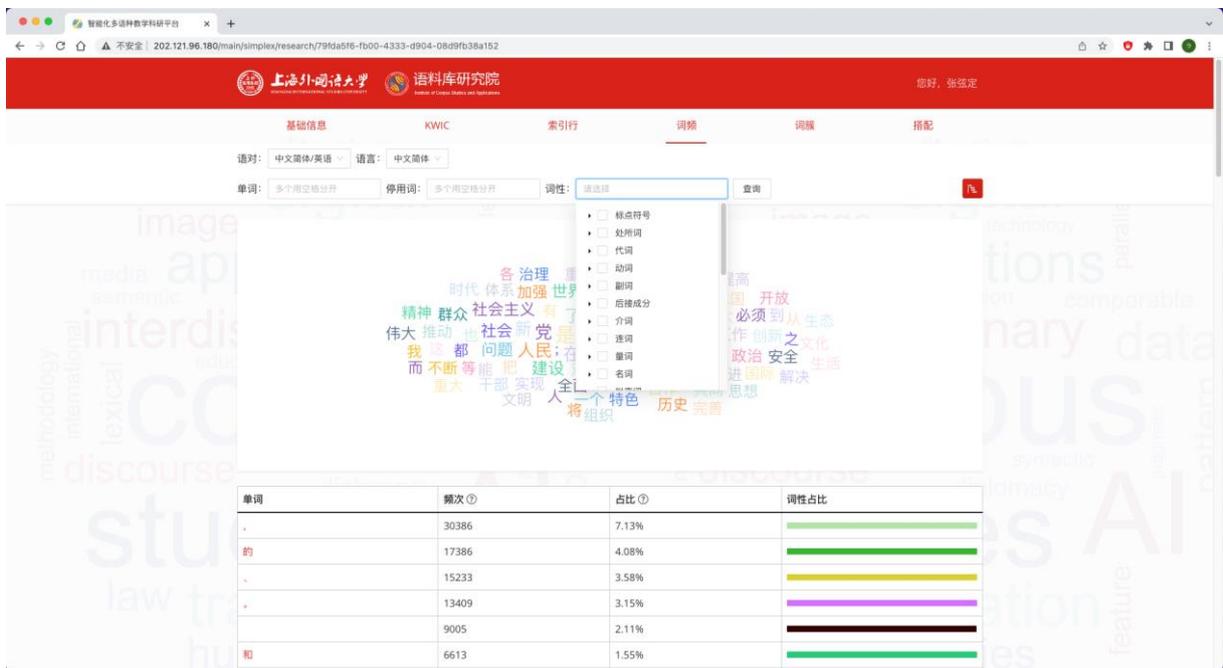


图 49

然后勾选需要统计的词性，此处以选择代词中的“代词性语素”、“人称代词”、副词中的“副词性惯用语”为例。

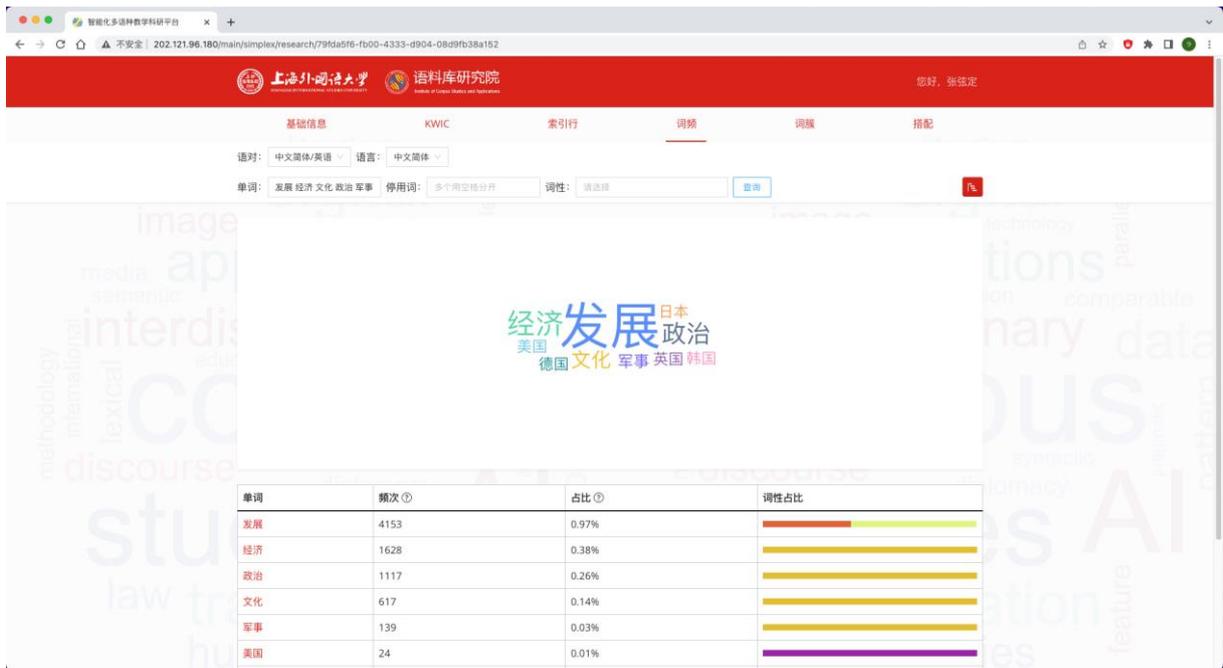


图 52

2.2.4.5 跳转至 KWIC 与索引行

在表格形式的词频信息中，用户可以点击其中的单词，平台可以自动跳转至对应单词在 KWIC 与索引行中的检索结果。

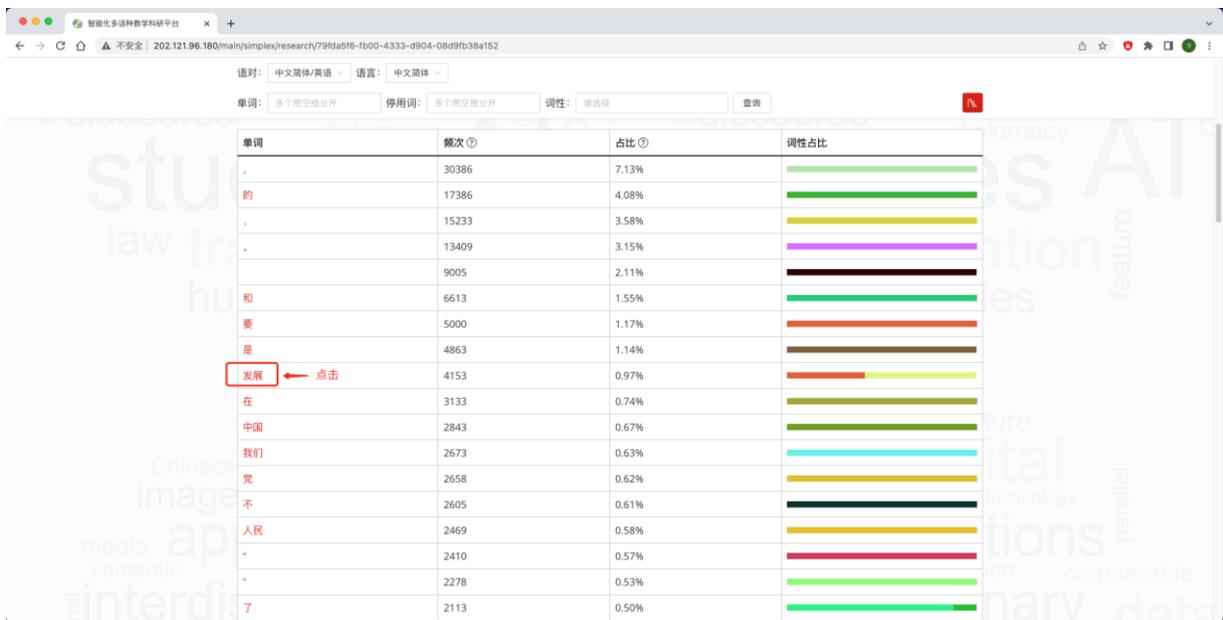


图 53

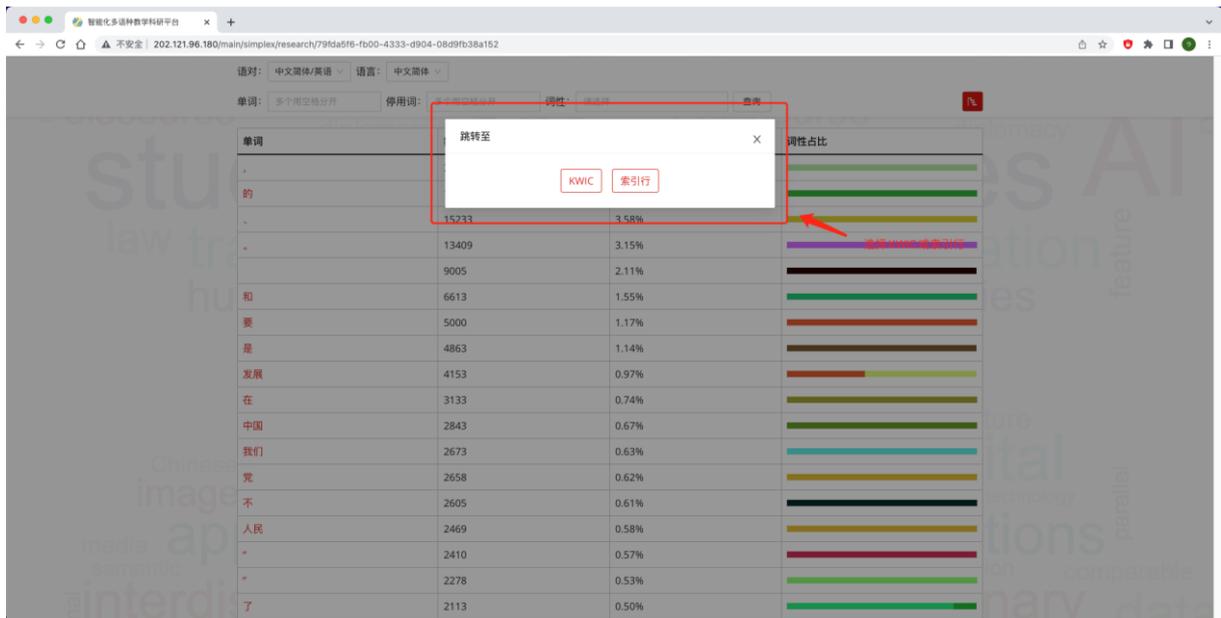


图 54

KWIC 检索结果如下：

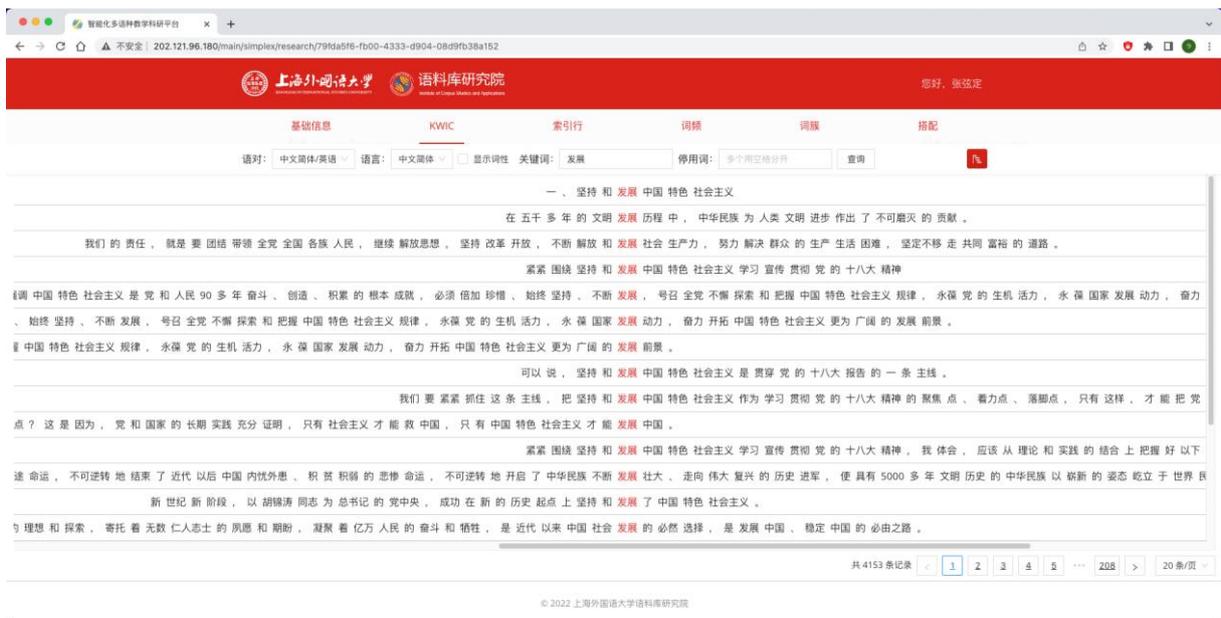


图 55

索引行检索结果如下：

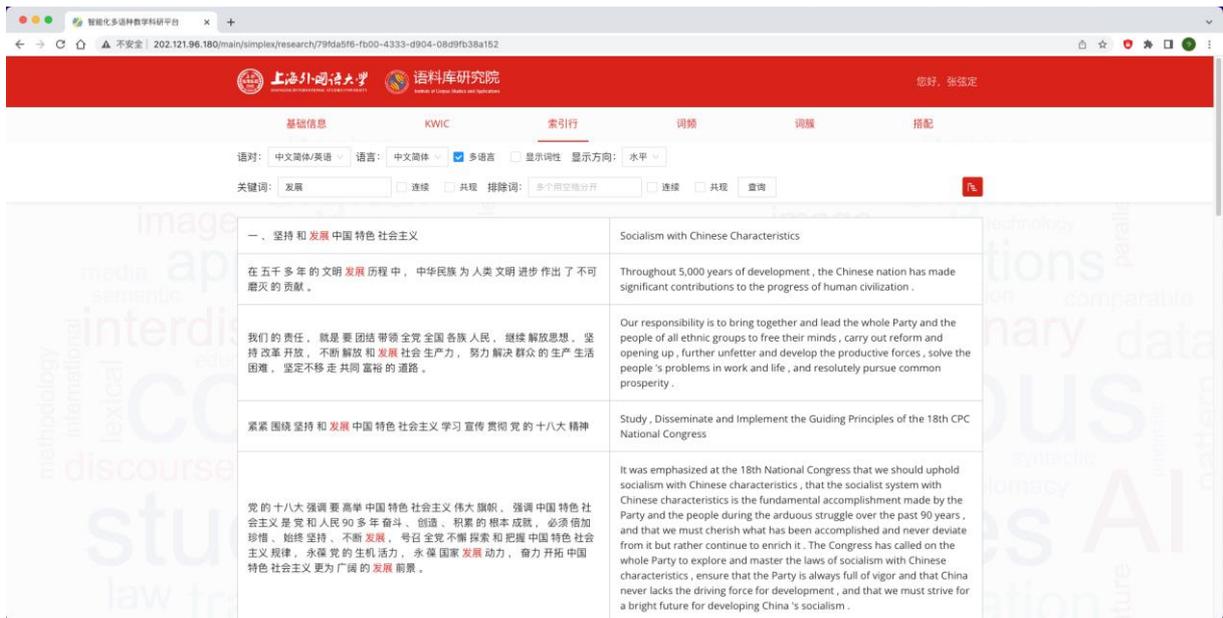


图 56

2.2.5 词簇

词簇功能可以呈现 2 个或 2 个以上的词构成的序列, 也可以输入检索词, 查询包含特定词汇的序列, 同时平台还会统计对应序列在选定的目录范围中出现的频次及其占比数据。

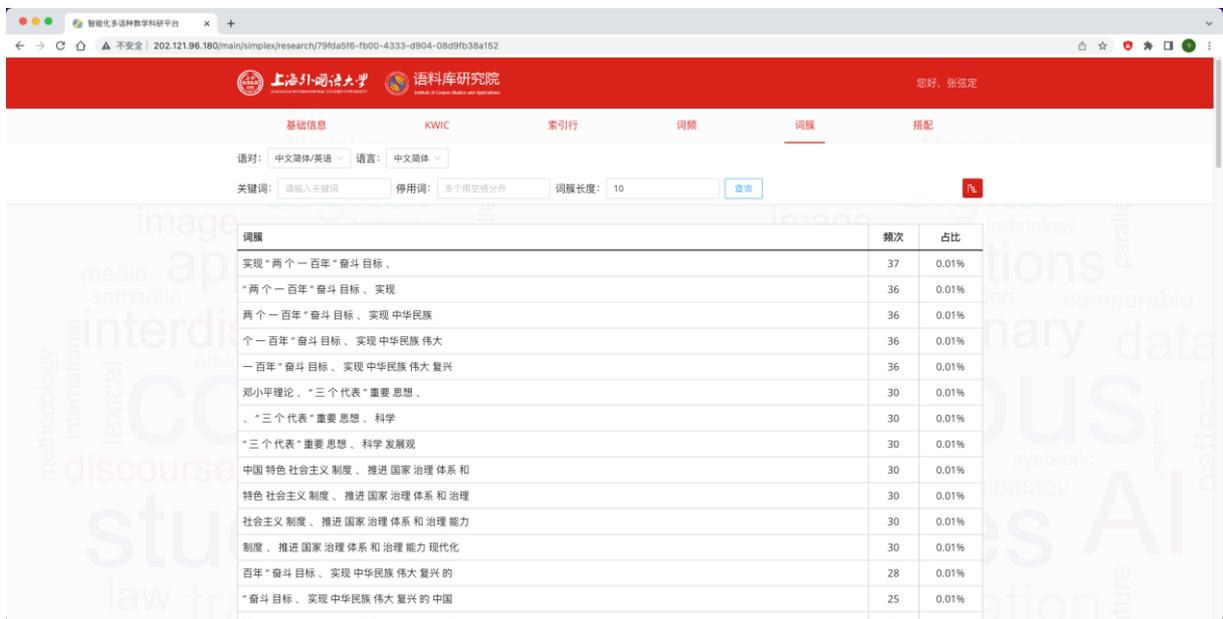


图 57

词簇	频次	占比
发展成果更多更公平惠及全体人民、	9	0.03%
、协调、绿色、开放、共享的发展	9	0.03%
协调、绿色、开放、共享的发展理念	9	0.03%
改革发展稳定、内政外交国防、治党	8	0.03%
发展稳定、内政外交国防、治党治国	8	0.03%
实现好、维护好、发展好最广大	8	0.03%
好、维护好、发展好最广大人民	8	0.03%
、维护好、发展好最广大人民根本	8	0.03%
维护好、发展好最广大人民根本利益	8	0.03%
完善和发展中国特色社会主义制度、推进国家	8	0.03%
和中国特色社会主义制度、推进国家治理	8	0.03%
发展中国特色社会主义制度、推进国家治理体系	8	0.03%
执政规律、社会主义建设规律、人类社会发展	7	0.02%
规律、社会主义建设规律、人类社会发展规律	7	0.02%

图 58

2.2.5.1 设置语对与语言及目录范围

具体操作请参照 [2.2.2.1 语对与语言](#) 及 [2.1.1.2 设置目录范围](#)。

2.2.5.2 关键词

用户可以通过输入关键词来检索包含关键词的词簇，以“发展”为例，检索结果如下：

词簇	频次	占比
发展成果更多更公平惠及全体人民、	9	0.03%
、协调、绿色、开放、共享的发展	9	0.03%
协调、绿色、开放、共享的发展理念	9	0.03%
改革发展稳定、内政外交国防、治党	8	0.03%
发展稳定、内政外交国防、治党治国	8	0.03%
实现好、维护好、发展好最广大	8	0.03%
好、维护好、发展好最广大人民	8	0.03%
、维护好、发展好最广大人民根本	8	0.03%
维护好、发展好最广大人民根本利益	8	0.03%
完善和发展中国特色社会主义制度、推进国家	8	0.03%
和中国特色社会主义制度、推进国家治理	8	0.03%
发展中国特色社会主义制度、推进国家治理体系	8	0.03%
执政规律、社会主义建设规律、人类社会发展	7	0.02%
规律、社会主义建设规律、人类社会发展规律	7	0.02%

图 59

2.2.5.3 停用词

用户可以设置停用词让平台在检索词簇时排除包含停用词的结果，以停用“人民”为例：

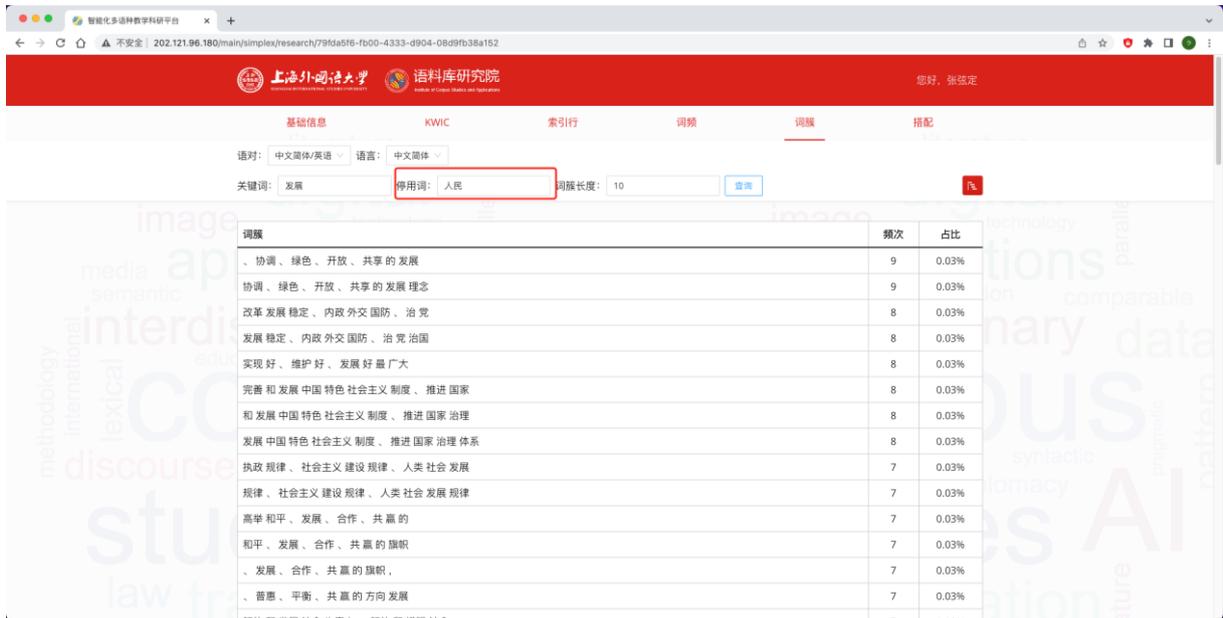


图 60

2.2.5.4 词簇长度

平台默认的词簇检索长度为 10，指的是检索长度为 10 个词以内词簇，用户可以根据自身需求对词簇长度进行调整，如将长度设置为 15：

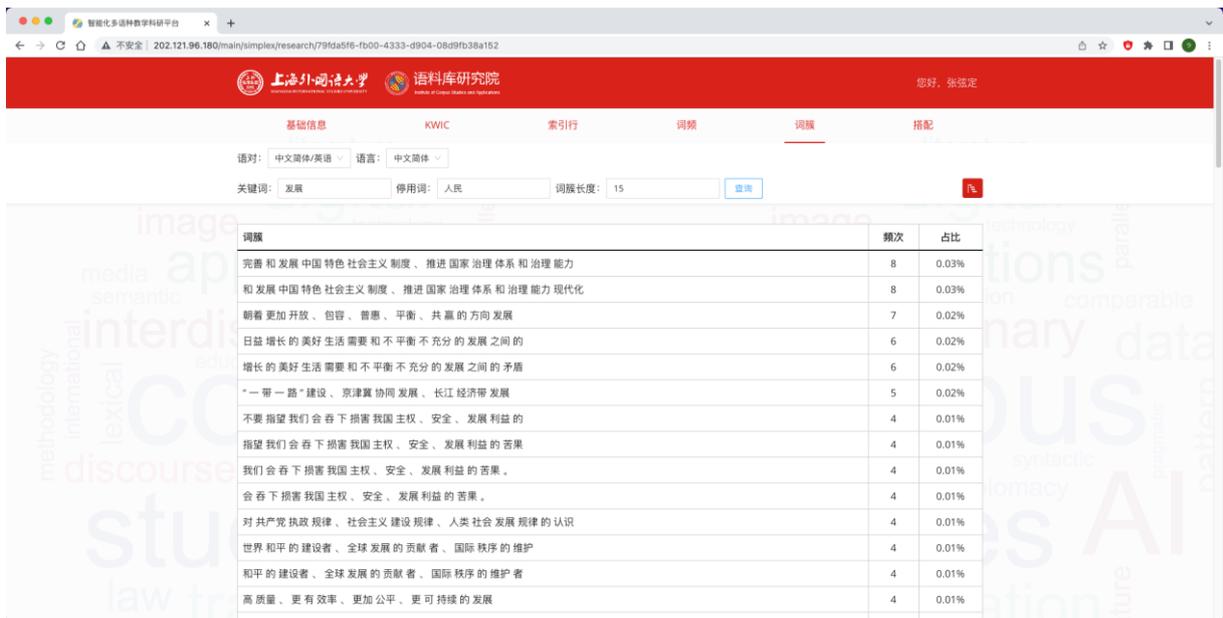


图 61

2.2.6 搭配

搭配（collocation）是指一些词与另一些词同时出现或共现的关系。

搭配模块可以用来考察词汇的共现。选定语种并输入要考察的节点词后，可以呈现与该节点词一起出现的其它词汇的信息，包括频次、搭配强度等信息。检索结果可以按搭配词的字母顺序、频率高低或搭配强弱等多种信息排序。在呈现具体搭配数据的同时，该模块也实现了搭配信息的可视化，直观地展示词语的搭配行为。

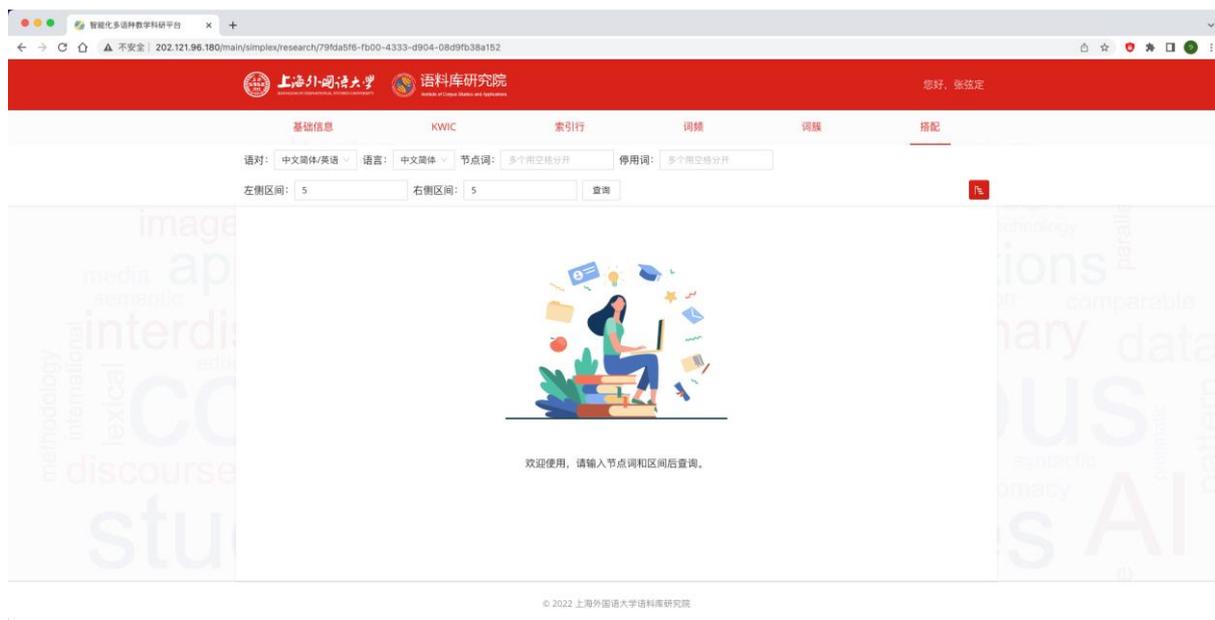


图 62

2.2.6.1 设置语对与语言及目录范围

具体操作请参照 [2.2.2.1 语对与语言](#) 及 [2.1.1.2 设置目录范围](#)。

2.2.6.2 节点词

进行搭配检索前，用户需确定作为搭配中心的节点词，以“中国”为例：



图 71



图 72

3.1 术语库子模块



图 73

在术语库子模块，用户可以输入关键词来检索与《习近平谈治国理政》相关的术语，并且术语库会提供该术语对应的英语译文，以“规划”为例，检索结果如下：

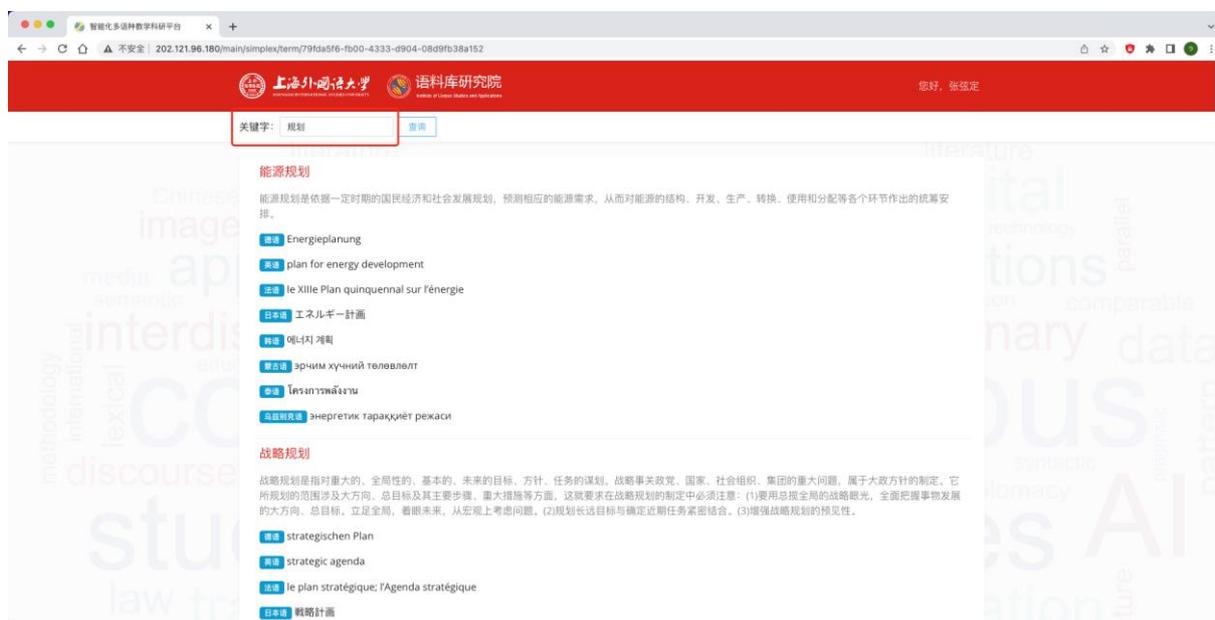


图 74

3.2 典故库子模块

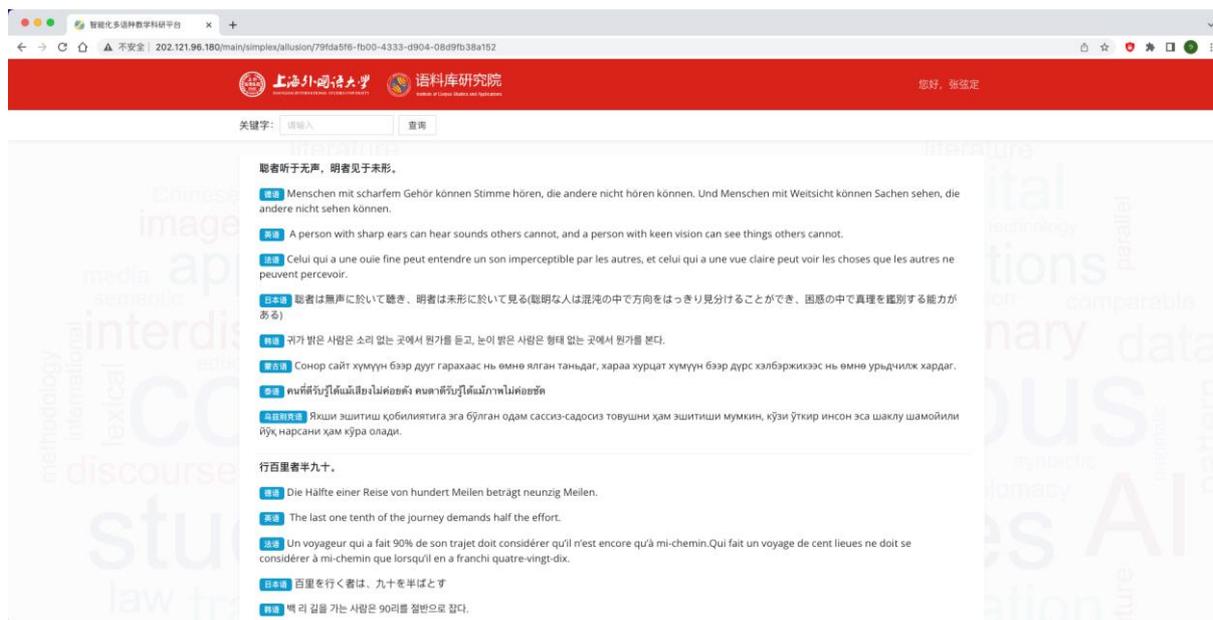


图 75

在典故库子模块，用户可以输入关键词来检索《习近平谈治国理政》中使用的典故，并且提供该典故对应的包括英语、法语、德语在内的 8 个语种的译文，以“德不孤”为例，检索结果如下：

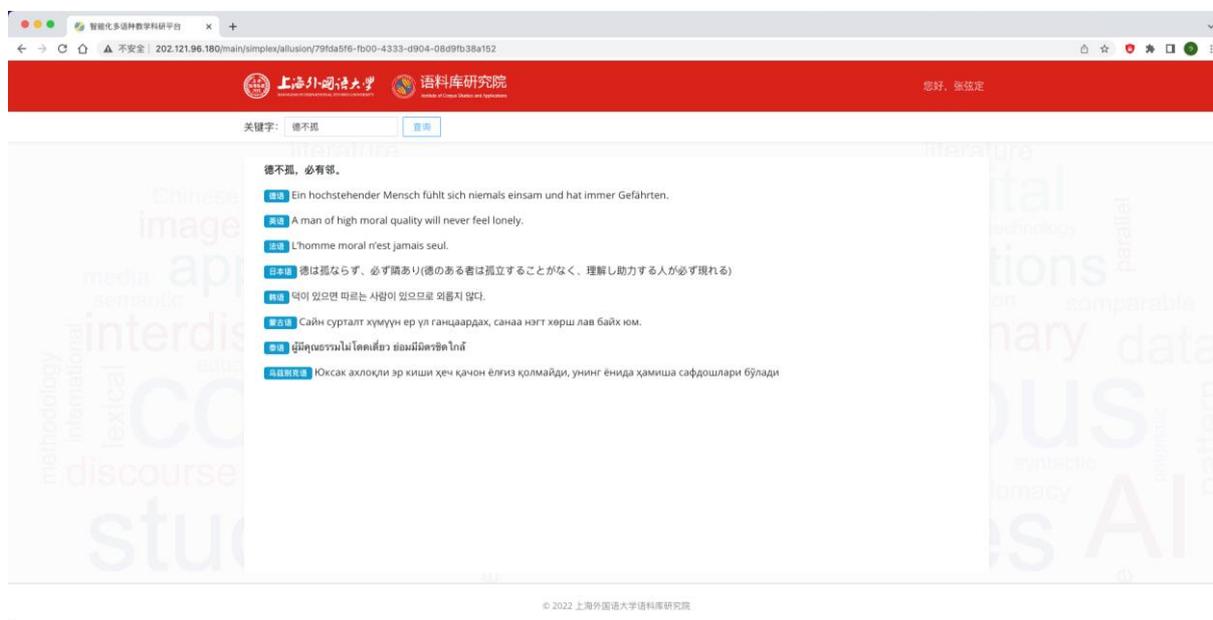


图 76

4. 文献库模块

文献库模块收录了与《习近平谈治国理政》相关的文本文献、图片文献及影像资料等。

在平台首页点击进入文献库模块：



图 77



图 78

进入文献库模块后，用户可以基于多种检索条件进行检索，如文献形式：



图 79

文献主题:



图 80

文献出处:



图 81

开始及结束日期：



图 82

文献语种：

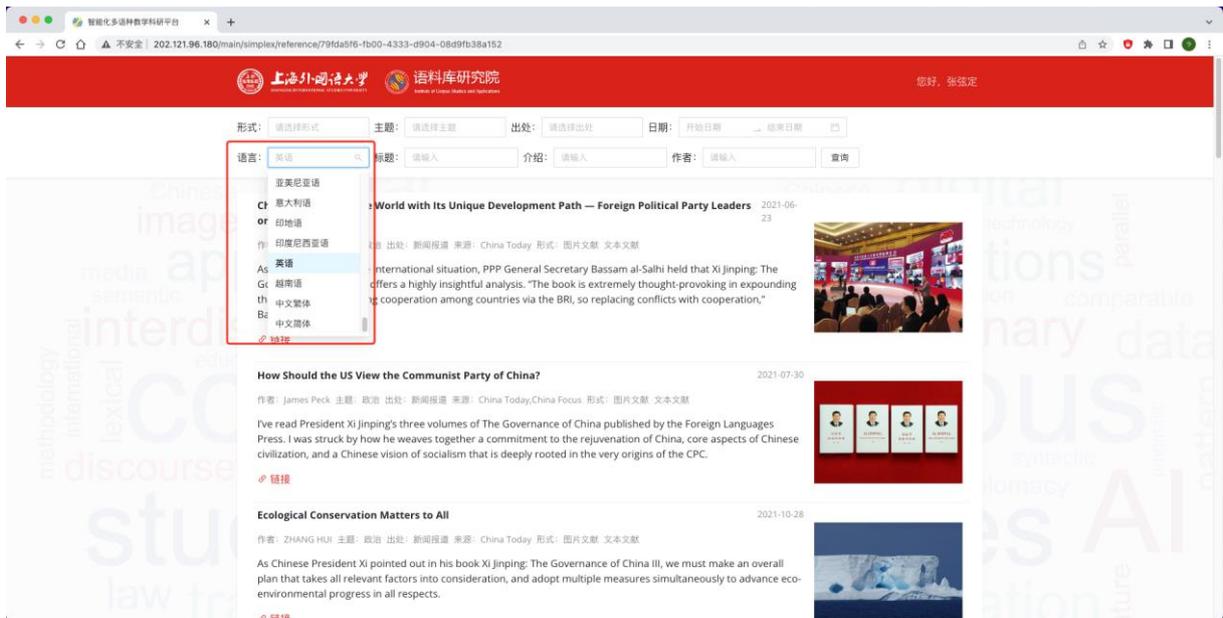


图 83

文献标题中的关键词：



图 84

文献介绍中的关键词：



图 85

文献作者:



图 86

在实际应用中，用户也可以结合多种检索条件进行复合检索，以更为精准的检索到符合要求的文献。

 上海外语教育出版社
外教社 SHANGHAI FOREIGN LANGUAGE EDUCATION PRESS

地址：上海市大连西路558号

电话：021-65425300

网址：www.sflep.com