

黄河·汾河流域诸方言数据库说明书

沈力·川崎广吉·郑弯弯·孙雪婷·星英仁
2026年3月31日

1. 数据库创建概要

黄河·汾河流域诸方言数据库是同志社大学语言生态科学研究中心跨学科研究成果的一部分。本研究旨在把握黄河流域及汾河流域中晋方言地区（北部）与中原官话地区（南部）之间所呈现的语言变化的阶段性，并通过提供基础数据以阐明两地区之间语言传播的实际状况。

本数据库是在日本学术振兴会科学研究费补助金·基础研究B（海外学术调查）“黄河流域方言混合地带语言传播实态解明——基于地理信息科学的方法——”（研究代表：沈力，课题编号15H05156，研究期间2015年~2018年）以及中国国家社会科学基金一般项目“基于GIS的汾河流域方言混合地带语言传播实态解明”（研究代表：沈力，课题编号18BY042，研究期间2018年~2021年）的资助下构建完成。

调查对象的区域（县、市）为属于晋方言与中原官话混合地带的以下地区：

（1）黄河沿岸（县以下的方言名称略）

- 陕西省一侧（西）：神木、佳县、绥德、吴堡、清涧、延川
- 山西省一侧（东）：兴县、临县、柳林、石楼、永和、大宁

（2）汾河沿岸（县以下的方言名称略）

- 汾河流域北部：文水、汾阳、孝义、祁县、平遥、介休
- 汾河流域中部：灵石、汾西、霍州
- 汾河流域南部：洪洞、临汾、稷山、新绛、侯马、襄汾、河津、万荣

本说明书首先叙述黄河·汾河沿岸诸方言数据库的构建经过，其后对本数据库的使用方法进行说明。

2. 方言数据库构建回顾

如上文（1）（2）所示，黄河流域的方言混合地带在陕西省一侧及山西省一侧各包含6个县市，汾河流域的方言混合地带在北部包含6个县市，中部（灵石高地）包含3个县市，南部包含8个县市，共计29个县市作为调查对象。最初预计方言区划数量与县市数量大致相当，但实际上方言类型非常多样。本研究所涉及的黄河流域方言共计51种（山西侧32种，陕西侧19种），此外汾河流域达到125种（北部58种，中部〔高地〕20种，南部47种）。

为阐明黄河·汾河流域的语言传播机制，构建方言数据库具有不可或缺的意义。本研究基于众多协作者的调查成果，对这些数据进行系统整理并汇编为数据库。

本数据库以①《汉语方言调查字表》（中国社会科学院语言研究所）为依据设定调查项目，共计3810字；②同时记录这些字对应的《广韵》音韵特征，③北京话中这3810字的声母、韵母和声调信息，总计达到11430项。此外，对于黄河沿岸的51种方言及汾河沿岸的125种方言，分别附加其声母、韵母、声调信息，并结合历史信息与现代语音信息一并收录。

数据库整体文字总量达到757539字。在本数据库中，例如检索声母时，可参照如下对应关系：

- 各方言中3810字的声母对应《广韵》的声母及北京话的声母
- 各方言中3810字的韵母对应《广韵》的韵母及北京话的韵母
- 各方言中3810字的声调对应《广韵》的声调及北京话的声调

通过利用本数据库，可对黄河秦晋沿岸·汾河流域南北地区诸方言的历史演变与地域变化进行系统性分析。

本研究实施周期为 2015 年度至 2021 年度，虽已按计划完成，但在数据库再次核验后，决定于 2025 年度通过“同志社大学语言生态科学研究中心”官方网站公开发布。通过对公开数据的利用，预期不仅能够分析黄河秦晋沿岸及汾河沿岸方言的音段与超音段特征，还能够进一步深化对语言变化实际状况的理解。

3. 数据库的构造和使用方法

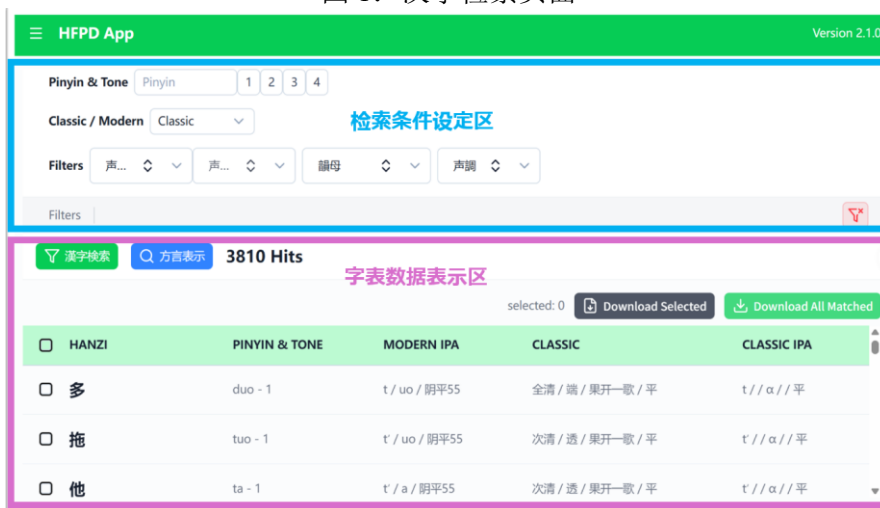
3.1 数据库的构造

黄河·汾河流域诸方言数据库由“汉字检索”和“方言检索”两大模块构成。

3.1.1 汉字检索

汉字检索模块由检索条件筛选区和字表数据展示区两部分组成。上方为检索条件筛选区，下方为符合条件的字表数据展示区。

图 1：汉字检索页面



(1) 检索条件筛选区的

该区域提供多种独立检索方式，主要包括：

- ①拼音及声调检索；
- ②现代汉语国际音标（IPA）标注检索；
- ③中古音条件检索。

各类检索方式相互独立，分别用于从不同层面对字表数据进行筛选。

(2) 字表数据区

该区域呈现符合检索条件的汉字条目。当前数据库共收录 3810 条字表数据，排序依据为《汉语方言调查字表》。每条数据依次提供：汉字、拼音标注、现代汉语国际音标表示、中古音对应形式及中古音构拟（参考王力体系）。字表上方提供检索结果的数据下载功能。

3.1.2 方言检索

方言检索模块由方言点筛选区和方言数据展示区两部分组成。

图 2：汉字检索页面



(1) 方言点筛选区

该区域用于对方言点进行分层筛选。一级筛选以流域范围为依据，用于区分两河流域、黄河流域和汾河流域；二级与三级筛选则在此基础上进一步结合行政单位信息，对方言点进行逐步细化和定位，从而实现目标方言点的有效筛选。

(2) 方言数据表示区

该区域用于展示与所选方言点对应的方言数据。数据库数据采集时间为 2015 年至 2018 年，共收录 75739 条记录。每条数据依次展示汉字、所属流域区、所属县名、方言名称以及方言的国际音标（IPA）标注。数据展示区提供检索结果的分页浏览与数据下载功能。

3.2 数据库的使用方法

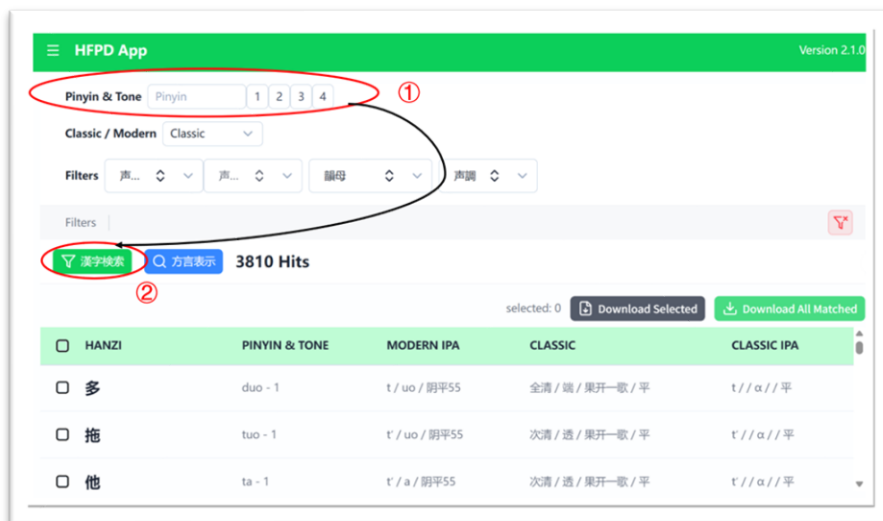
3.2.1 汉字检索

(1) 拼音声调检索

数字 1~4 分别代表的是现代汉语普通话的第一声~第四声。在空栏处输入想要检索的汉字的拼音，选择相应声调后，点击位于字表左上方的“汉字检索”，则字表区就会出现筛选结果。上述操作顺序可参考图 3。

关于声调的选择，可以多选，点击声调数字标号后，按钮变黑表示选中成功，变黑后再点击一次，按钮变成白色可以取消选择。

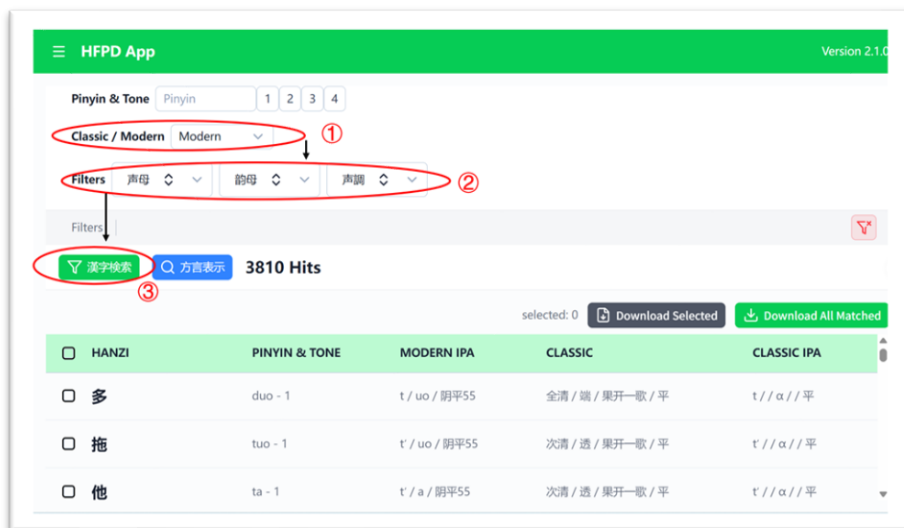
图 3：拼音声调检索图例



(2) 现代汉语国际音标(IPA)检索

在检索条件筛选区选择“Modern”，此时“Filters”会出现声母、韵母、声调三个选择下拉菜单，选择目标汉字的国际音标表示即可。三个选项的选择是任意自由的，可以单选，也可以多选。选择完毕后点击下方的“汉字检索”按钮，字表区就会出现符合条件的汉字。上述操作顺序可参考图 4。

图 4：现代汉语国际音标(IPA)检索图例

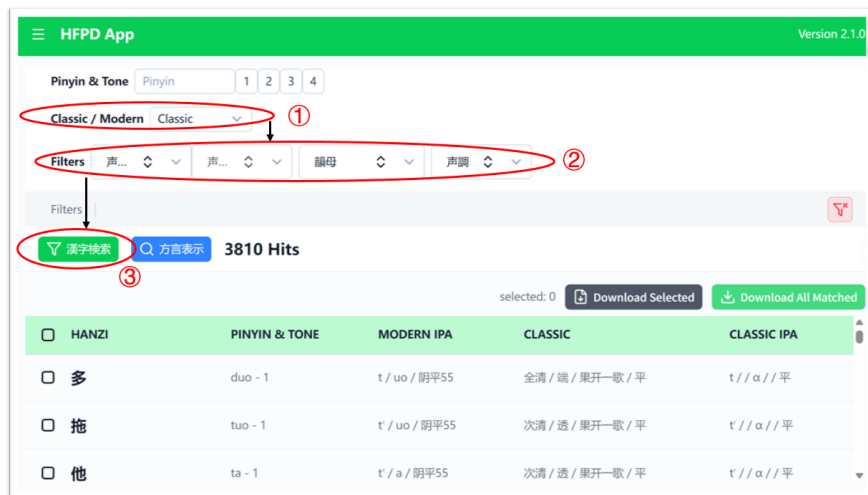


(3) 中古音检索

在检索条件筛选区选择“Classic”，此时“Filters”会出现声母 1、声母 2、韵母、声调四个选择下拉菜单。声母 1 是对声母清浊性质的选择，声母 2 是具体的古声母字名称的选择。声母 2 的内容多，为了方便声母 2 的选择，在这里我们将声母 1 和声母 2 之间设定成了有制约关系的选择操作，即选声母 2 前必须先选择声母 1。此外，为了满足多种筛选要求，以下选择方式都可以进行操作：可以单选声母 1 不选声母 2；韵

母和声调的选择是自由的，单选韵母或声调也可以对汉字进行筛选。上述操作顺序可参考图 5。

图 5：现代汉语国际音标(IPA)检索图例

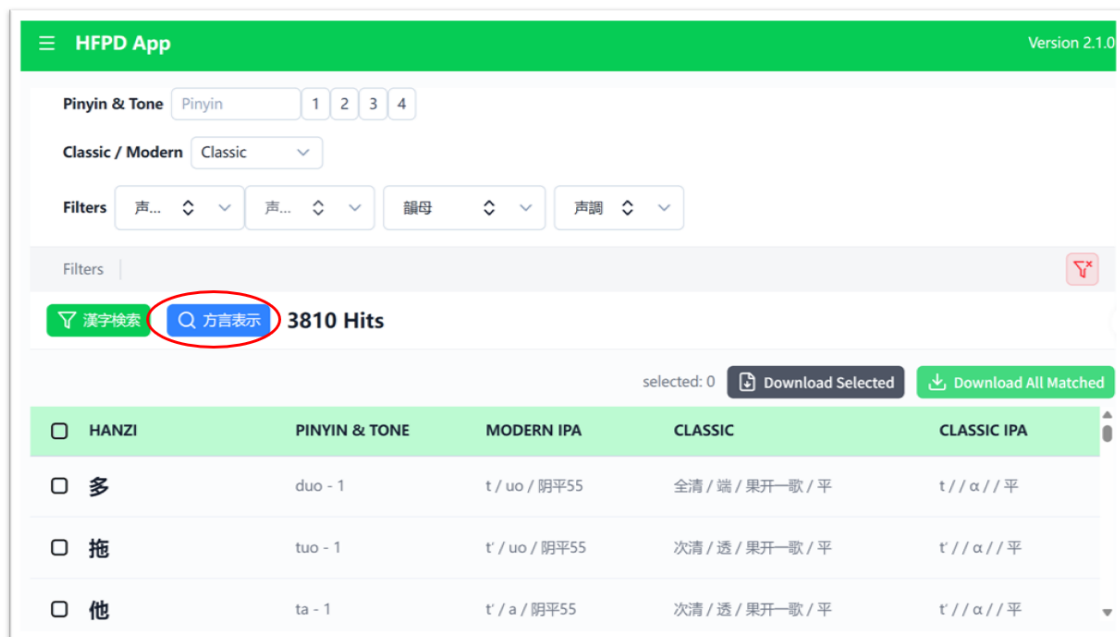


3.2.2 方言检索

(1) 由“汉字检索”页面进入“方言检索”页面

符合要求的汉字检索完毕之后，点击“方言表示”就可进入方言数据页面。此时默认的是筛选出来的所有汉字作为查看内容，进入方言数据页面。操作按钮如图 6 所示。

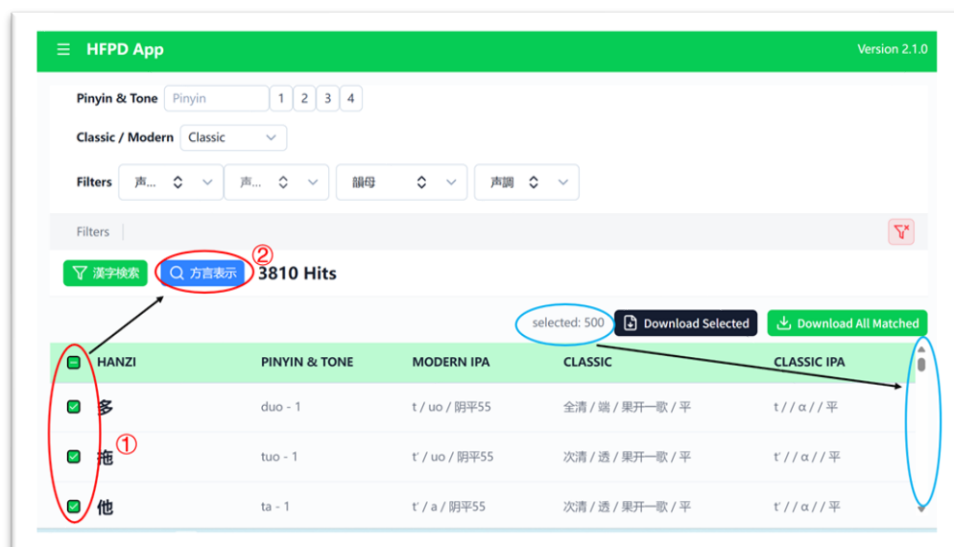
图 6：由汉字检索进入方言检索的操作图例



如果相对筛选出来的汉字再进一步进行选择后查看方言数据，可以在字表区域中，汉字列的左侧方框内进行勾选。可以单独勾选目标汉字，也可以批量勾选。批量勾选的时

候点击“HANZI”左侧的方框，一次勾选的默认值为 500 个汉字，如果在这 500 个汉字的基础上还想增加，可通过下拉右侧滑动条的方式使字表数据进一步更新后，再点击“HANZI”左侧的方框。具体操作方式的示意图可参看图 7。

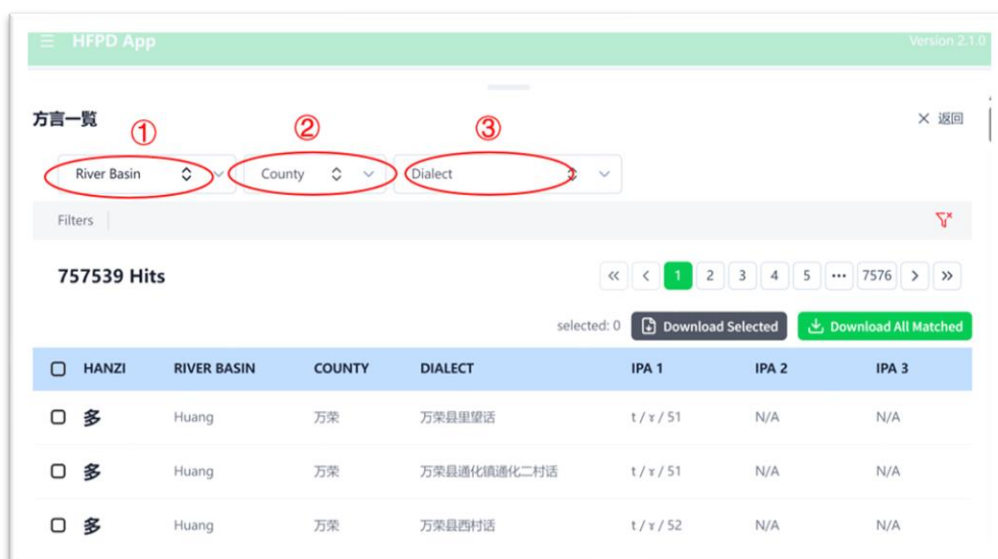
图 7：对汉字进行二次筛选后进入方言数据页面的操作图例



(2) 方言点的筛选

筛选的项目共有三项，因为方言点过多，为方便高效检索，需要按照从左向右的顺序逐步精确。检索条件设置完毕后，按键盘上的“Enter”键，下方的字表区就会出现筛选结果。三个项目的位置说明可参考图 8。汾河和黄河的数据可以混合在一起检索。

图 8：方言点筛选按钮示意图例



3.3 其他补充说明

(1) 下载功能

汉字检索页面和方言检索页面，字表区域的右上方都有两个下载数据的按钮。如果只想下载筛选结果的一部分符合研究需要的信息可以在对汉字左侧的方框进行勾选后，点击“Download selected”按钮。如果需要下载所有筛选结果可以点击“Download All Matched”按钮。

(2) 查看汉字注释功能

但凡有注释的汉字，汉字右侧都会有绿点标记，将鼠标放在绿点上，注释就会浮现出来。

(3) 翻页功能

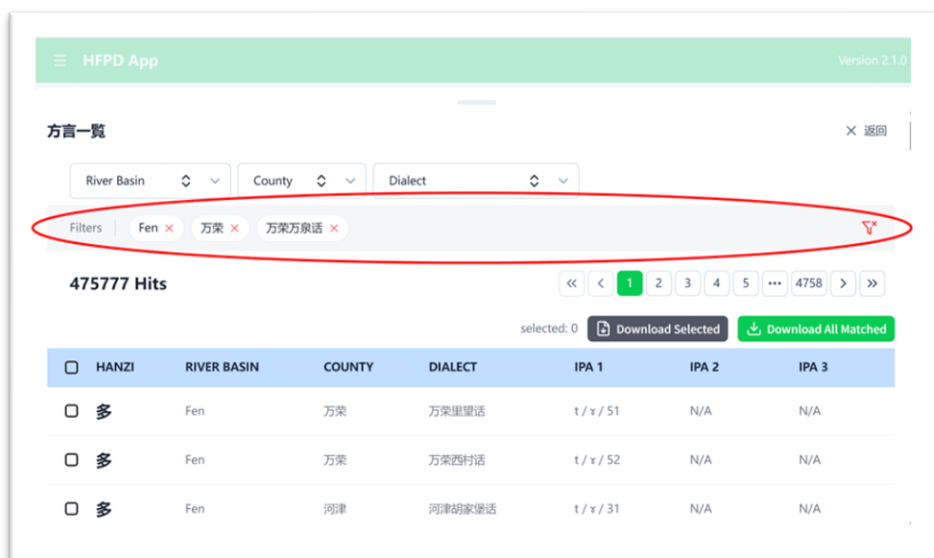
汉字检索页面的翻页功能可以通过向下拉动字表区右侧的滑动条实现。

方言检索页面的翻译功能可以通过点击字表区右上方的分页导航器实现。

(4) 筛选条件区的说明

汉字检索页面和方言检索页面，在筛选区下方的灰色区域都会实时出现已选择的筛选条件。点击每一个筛选条件右上方的红叉号，可以单独取消某一项的筛选条件。点击灰色区域最右端的红色漏斗，可以一键取消所有已选择的筛选条件。

图 9: Filters 筛选区的操作示意图例



(5) 从“方言检索”页面返回“汉字检索”页面

点击方言检索页面右上角的“返回”按钮，可以返回到汉字检索页面。切勿点击整个页面左上角的返回箭头，这样会回到默认浏览器的主页，失去刚刚筛选出来的数据信息。

图 10：从“方言检索”页面返回“汉字检索”页面的图例

