



库、学、研三位一体： 中国英语学习者语料库研究新思路

许家金

北京外国语大学
中国外语与教育研究中心



大数据视野下的学习者语料库研究



要点

- 语言大数据的核心特征
- 库、学、研三位一体的iWrite中国英语学习者语料库



大数据视野下的中国英语学习者语料库

- iWrite Corpus
- Size: 项目结束时，达到2亿词次
长期目标：动态监控语料库
- Structure: 丰富的元信息（XML demo）
- Speed: 利用服务器集群处理数据

<Affiliation name="学校类型"></Affiliation>
<School_at_universtity name="学院/系"></School_at_unive
<Grade name="年级"></Grade>
<Class name="班级"></Class>
<Exam_type name="使用场景"></Exam_type>
<Essay_translation name="语料类型"></Essay_translation>
<Prompt name="作文标题">IS STRESS A BAD THING</Prom
<Instructions name="作文指令"></Instructions>
<Student_start name="写作开始时间"></Student_start>
<Submission_time name="写作结束时间"></Submission_tin
<Exam_start name="考试开始时间"></Exam_start>
<Exam_end name="考试结束时间"></Exam_end>
<Total_points name="试题总分"></Total_points>
<Score name="人评分数"></Score>
<E-score name="机评分数">



iWrite Corpus/iWrite语料库

- 监控库：iWrite Corpus
- 精品库：iWriteBaby 1千万-2千万平衡库
- 定制库：iWriteCustom 按学校类型、
话题、体裁等随时生成



iWrite语料库基本理念

- 库学同源
- 库研同步
- 库教同理

- 将“教”、“学”、“研”三者融为一体
- 产品和过程兼顾



iWrite语料库基本理念

- 库学同源
- 库研同步
- 库教同理

- 将“学”、“教”、“研”三者融为一体
- 产品和过程兼顾



- **“库学同源”** 的观念认为，学生练习写作之时即是学习者语料库建成之日。
- **“库研同步”** 的思路是指建库的同时，教研人员可以实时对学生进行诊断性语料分析，并给出反馈。
- **“库教同理”** 的观念是指语料库的建设与研究 and 基于学生作文语料的教学活动，均是遵循数据驱动的方法。



库₁、库₂、库₃

- 大数据语料库₁：免费语料库平台
- 学生作文错误库₂：写作现状皮书
- 教学案例库₃：分享交流平台



学习者语料库检索技术新进展

- 错误检索
- 错误搭配模式分析



在下方选择统计条件，然后点击统计按钮

Size : 76199210 Millon

Subcorpus-selected : 76199210Million

时间

3

全部

2015

2016

2017

作文分数

6

全部

100-90

89-80

79-70

69-60



统计结果

词法类	468716
-----	--------

介词短语错误	589
--------	-----

代词格错误	16778
-------	-------

冠词错误	51800
------	-------

动词及物-不及物	20
----------	----

动词时态错误	3633
--------	------

动词短语错误	743
--------	-----

句法类	789707
-----	--------

动词词类	84551
------	-------

存现句[there_be]错误	14941
-----------------	-------

定语从句关系词错误	45272
-----------	-------

流水句	53315
-----	-------

语序错误	485
------	-----

谓语动词缺失	11149
--------	-------

连词缺失	42443
------	-------

搭配类	4798
-----	------

惯用语错误	4798
-------	------

技术规范	1388897
------	---------

拼写错误	1385346
------	---------



4 Secondly , there are some student **think** the studying is not important , playing than studying .

5 Secondly , there are some **student** think the studying is not important , playing than studying .

6 Reviewing my first year in university , I am taking my college life as a str things for me to worry about .

7 In the library there **have** many books .

8 If there are some tasks **remained** to me , I will finish them perfectly .

9 There are many companies **need** mostly person who has strong creator

10 After I got it , there are many changes **happened** during my life .



7 One day , your world out of the question , but found that n
can go to a place , not so much , is the world's abandoned mys
, rather , isyou **abandon** the whole world .

8 Thank you , they gave me life , thank you for taking care of
me they have never **abandon** .

9 We **shuould abandon** the use of computers .

10 The day I met a sweeps in the elevator wearing plain , I
looked at her with **abandon** eyes and think she is ugly and dirt
also put my hand over my mouth and nose .



Thank you!
