

学校编码: 10384

分类号\_\_\_\_\_密级\_\_\_\_\_

学号: 12020101152634

UDC\_\_\_\_\_

基于语料库的高中英语写作错误分析

# 厦门大学

## 硕士学位论文

### 基于语料库的高中英语写作错误分析

#### A Corpus-based Analysis of Errors in High School Learners' Writings

吴凡

吴凡

指导教师: 李力教授

指导教师姓名: 李力教授

专业名称: 英语语言文学

论文提交日期: 2013年 月

论文答辩日期: 2013年 月

学位授予日期: 2013年 月

厦门大学

答辩委员会主席: \_\_\_\_\_

评阅人: \_\_\_\_\_

2013年 月



## 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为( )课题(组)的研究成果,获得( )课题(组)经费或实验室的资助,在( )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日



## 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1.经厦门大学保密委员会审查核定的保密学位论文，  
于 年 月 日解密，解密后适用上述授权。

2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日



## Abstract

Errors contain useful information on the strategies that learners employ to acquire a language and to communicate in the language. As the study on errors increases, the paradigm of Error Analysis (EA) comes into being. It is the process of defining the error, its features, causes and possible solutions. The analysis of errors is beneficial for researchers to find the causes and solutions, for teachers to resolve learners' learning difficulties, and for learners to be aware of the errors they are likely to make.

There are a number of domestic research on learners' errors in the process of SLA, most of which focus on college students' writings. This thesis, however, studies errors in high school students' writings. It is a corpus-based study consisting of two samples, each with 40 compositions from the 3<sup>rd</sup> year students in a key high school of Heilongjiang Province. Errors are classified at three levels, i.e. substance, text and discourse, each of which are further categorized into different types. The causes, including the interlingual, intralingual and communicative strategy based causes, are discussed. The two samples are compared and the results show that the improved language proficiency changes the distribution of errors, reduce the total numbers of errors and diminishes the influence of all three causes on these errors.

There are also a few limitations. The uncontrollable nature makes it impossible to adjust the materials or activities to get the intended results. Some errors are corpus-specific, which are only applicable within the samples. It is hard to determine the dominant cause of a few types of errors with more than one cause. The suggestions for improvement would be to involve students in giving feedbacks that provide their reasons, to study the non-linguistic aspects such as MT thinking and metacognitive strategies in writing, to establish a corpus that includes past papers of the College Entrance Examination (the Exam) and to incorporate the native language corpora in designing the questions and marking guide for the Exam.

**Key Words:** Errors; Corpus; High school writings; Language proficiency

## 摘要

错误能提供有关学习者在语言学习及交流时使用策略的有效信息。随着对错误研究的增加，错误分析原理应运而生。它涵盖错误的分类、特点、原因及解决途径等内容。错误分析有以下三方面作用：它有助于研究者找出产生错误的原因及解决方法，有助于教师了解学生可能使用的学习策略并帮助他们解决所遇到的困难，有助于学习者认识自己易犯的错误并尝试避免。

国内已有许多关于二语习得过程中所犯错误的研究，其中大多数关注于大学生写作。本文则主要关注高中学生的写作错误。本次研究所采用的语料库包含两个样本，各有四十篇来自黑龙江一所重点高中高三学生的作文。错误被分为三个层次：本体错误，文本错误以及语篇错误，每个层次中还有进一步的划分。同时，也将对语际、语内以及交际策略等原因进行讨论。通过比较两个样本的结果表明：语言水平的提高会改变错误的分布、降低错误的数量以及减弱各种原因对于错误的影响。

本次研究也存在一些问题。研究的不可控性使我们无法通过调整材料及活动来达到预期的效果。一些错误的分类具有局限性，仅适用于本次所使用的语料库。还有一些错误由多种原因引起，但其中的决定性因素难以判断。本文对于未来的研究也提出了一些构想：邀请被试参与，提供错误原因；扩展研究范围，涵盖非语言因素如母语思维及元认知策略；建立高考语料库，结合使用本族语语料库用于高考试题以及评分标准设计。

**关键词：**错误 语料库 高中写作 语言水平



## Table of Contents

<b>Abstract.....</b>	<b>I</b>
<b>Chapter 1 Introduction.....</b>	<b>1</b>
1.1 Data and Methodology .....	1
1.2 Purpose of the study .....	2
1.3 Organization of chapters .....	2
<b>Chapter 2 Literature Review .....</b>	<b>3</b>
2.1 Definitions of errors .....	3
2.2 Paradigms on errors.....	3
2.2.1 Contrastive Analysis .....	5
2.2.2 Error Analysis .....	6
2.2.3 Interlanguage.....	7
2.3 Error Analysis.....	9
2.3.1 Taxonomies of errors .....	9
2.3.2 Recent studies on Error Analysis .....	12
2.2.3 Significance of this study.....	15
<b>Chapter 3 Classification of Errors .....</b>	<b>16</b>
3.1 Substance errors .....	17
3.1.1 Punctuation errors .....	17
3.1.2 Spelling errors .....	18
3.2 Text errors .....	19
3.2.1 Lexical errors .....	19
3.2.2 Grammar Errors .....	23
3.3 Discourse errors.....	31
3.3.1 Coherence .....	31
3.3.2 Pragmatics.....	32

<b>Chapter 4</b>	<b>Diagnosis of Errors .....</b>	<b>34</b>
4.1	Interlingual causes.....	34
4.2	Intralingual causes .....	36
4.3	Communicative strategy based causes.....	37
4.3.1	Holistic strategy .....	37
4.3.2	Analytic strategies .....	38
<b>Chapter 5</b>	<b>Comparisons and Discussions .....</b>	<b>39</b>
5.1	Interlingual causes.....	41
5.2	Intralingual causes .....	42
5.3	Communicative strategy based causes.....	43
5.4	A Brief Summary .....	44
<b>Chapter 6</b>	<b>Conclusion .....</b>	<b>45</b>
6.1	Limitations of the study .....	45
6.2	Suggestions for further study .....	46
<b>References</b>	<b>.....</b>	<b>49</b>
<b>Acknowledgement</b>	<b>.....</b>	<b>54</b>

廈門大學博碩

# 目 录

摘 要.....	II
第一章 绪论 .....	1
1.1 研究数据及方法 .....	1
1.2 研究目的 .....	2
1.3 篇章结构 .....	2
第二章 文献综述 .....	3
2.1 错误的定义 .....	3
2.2 错误的研究范式 .....	3
2.2.1 对比分析 .....	5
2.2.2 错误分析 .....	6
2.2.3 中介语理论 .....	7
2.3 错误分析 .....	9
2.3.1 错误的分类 .....	9
2.3.2 近期对错误分析的研究 .....	12
2.2.3 本研究的意义 .....	15
第三章 错误层次 .....	16
3.1 本体错误 .....	17
3.1.1 标点错误 .....	17
3.1.2 拼写错误 .....	18
3.2 文本错误 .....	19
3.2.1 词汇错误 .....	19
3.2.2 语法错误 .....	23
3.3 语篇错误 .....	31
3.3.1 连贯 .....	31
3.3.2 语用 .....	32

第四章 错误诊断 .....	34
4.1 语际错误 .....	34
4.2 语内错误 .....	36
4.3 交际策略引起的错误 .....	37
4.3.1 整体策略 .....	37
4.3.2 分析策略 .....	38
第五章 对比及讨论 .....	39
5.1 语际对比 .....	41
5.2 语内对比 .....	42
5.3 交际策略对比 .....	43
5.4 小结 .....	44
第六章 结论 .....	45
6.1 本研究的局限 .....	45
6.2 对后续研究的建议 .....	46
参考文献 .....	49
致谢 .....	54

## Chapter 1 Introduction

Error is defined as the “deviation of the norms of TL” (Ellis, 1994:51). Learners’ errors had been considered something undesirable and should be avoided. The fear of errors did not change until the end of 1960s, when Corder (1967) stated that errors could be significant, and EA then became a paradigm of applied linguistics. The analysis of errors is useful in that it reveals the conspicuous features of learner language, it tells teachers what errors learners are likely to make and it helps learners to improve their learning when they correct the errors they make (Ellis, 1997).

Quite a few studies have been conducted on foreign language learners’ errors in writing, most of which take college students as the subjects (马广惠, 2001; 刘东虹, 2004; 黄莹、陈建平, 2006; 王立非、张岩, 2007; 韩松, 2008). However, there are relatively fewer studies on high school students’ writings. As one of the four basic skills that takes up to 25% of the overall scores in the College Entrance Examination, writing is of great importance in high school English teaching and learning. The third year of the high school study is crucial in that the systemic review on grammar and intensive training on language usages help to enhance learners’ language proficiency, which also lead to the changes in learners’ errors.

### 1.1 Data and Methodology

This thesis is a corpus-based study that focuses on errors in high school students’ writing. The subjects are 40 students in their 3<sup>rd</sup> year from a key high school of Heilongjiang Province. The corpus consists of two samples, each with 40 writings composed by the same subjects. The data in Sample 1 is from the 1<sup>st</sup> test at the beginning of their 3<sup>rd</sup> year study in September 2012, and the data of Sample 2 is from the last test at the end of their study period in June 2013. The data in both samples are compositions with given topics and a few pieces of information is provided after each topic. The topic in Sample 1 is to describe the NBA star - Lin-sanity and Sample 2’s

topic is to introduce a host family to an American student who joined the summer camp. The requirement is to complete a composition within 80-120 words and the marking scale ranges from 0-25.

## **1.2 Purpose of the study**

The study is to provide a reference for the classification and causes of these errors, to explore the relations between changes in learners' language proficiency and changes in their errors, and to seek solutions to reduce the errors and improve learners' writings. Moreover, it is to address for the need of more studies on the errors of high school students' writings.

## **1.3 Organization of chapters**

The proceeding part is divided into four parts. Chapter 2 is Literature Review which is a brief overview of three areas of language research, i.e. Contrastive Analysis, Error Analysis and Interlanguage, each of which dominated for some time and constitute the current research paradigms and new directions in second language studies. Chapter 3 is the categorization of errors at three levels, namely, substance, text and discourse, some of which are with subcategories. Examples will be taken from Sample 1 to analyze each type of error. Chapter 4 is the diagnosis of the errors which includes interlingual (mother-tongue interference), intralingual and communicative strategy based causes. Chapter 5 is the comparison of Sample 1 with Sample 2 to study the impact of increased language proficiency on the distribution and causes of these errors in Sample 2. Reasons for the change and implications from the change will also be discussed. Chapter 6 is to summarize the limitations of this research and provide suggestions for further studies.

## Chapter 2 Literature Review

### 2.1 Definitions of errors

We are now in a position to construct errors. There are definitions on errors from different perspectives. Corder (1967:167) associates errors with competence and upholds that errors reflect learners' knowledge and are not self-correctable. Later, Corder (1971:152) shifts ground and states that errors result from performance failure and they are readily correctable. The basic criterion in Corder's definition is self-correctability. However, Corder does not explain to what extent these errors are self-correctable. Edge (1989) makes it clear that errors are wrong forms that learners are unable to correct even if their wrongness were pointed out. Unlike Corder (1967,1971) and Edge(1989) that applies the test of self-correctability, Hammerly (1991:72) focuses more on the classroom or formal learning context and he (ibid.) claims that "faults" (the term used to replace errors) occur when learners make attempts to express their ideas freely but err on the structures that they have not been taught.

In this thesis, learners' errors are defined as unsuccessful bits when using language. No appeal is made to self-correctability and learners err on what they have already been taught.

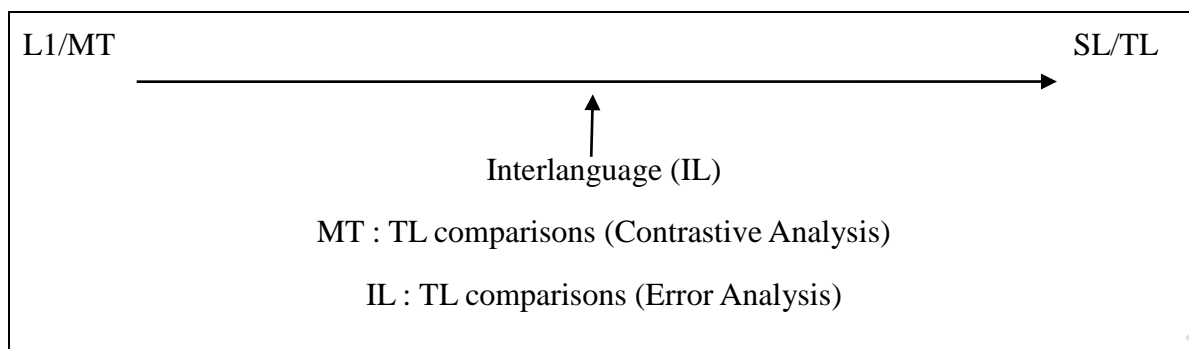
### 2.2 Paradigms on errors

After the definition, we will now move to the methodologies that are used to study errors. There are three paradigms which involve three language codes, and we will start from the discussion of these language codes.

First, it is the description of the target language (TL) to be learnt. It is the target that learners are striving to achieve and it can be extended to include the goals the learners set for themselves.

Second, it is the description of learners' version of TL or as defined by Selinker (1972), their Interlanguage (IL). It is called "idiosyncratic dialect" by Corder (1971) that is of TL standard but is different from the TL. Another related label is the "whole performance data from individual learners" (Corder, 1975:207). It is further developed by Faerch (1978) as "performance analysis" that studies learners' (with the same L1) TL-oriented repertoire.

Third, it is the description of learners' MT. It should be at least one more language involved in the learning process. For people who have known more languages, MT should be "the totality of the TL learners' prior linguistic knowledge" with diversified degrees of understanding (James, 2001:4).



**Figure 2.1** Points of comparison for the paradigms (adapted from James, 2001:3)

The comparison of three language codes, i.e. MT, TL and IL has yielded two paradigms (Figure 2.1), i.e. Contrastive Analysis (CA) that compares MT with TL and Error Analysis (EA) that compares IL with TL. The third widely used paradigm is the Interlanguage hypothesis produced by Selinker (1972). The following section is to discuss the three paradigms respectively. The three paradigms are not contradictory but complementary (Sridhar, 1976) with each other and they are viewed at three phases of the same goal, which is to deal with problems with learning difficulty and probes into the nature of second language learning process.

Contrastive Analysis assumes that the difference between native and target languages are the causes of many mistakes made by learners. It has led to considerable amount of valuable language description materials and pedagogical grammars. Error analysis provides evidence of second language learning process by



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

廈門大學博碩