## 概述

当前主流的语法检查主要分为两类：基于规则和基于实例的语法检查。在之前的工作中，我们调研了现有的大部分开源语法检查器，并对部分效果较好的语法检查器进行了仔细的分析和研究，在借鉴各检查器的主要思想后，我们对这些检查器进行了改进及重构，期望得到更好的效果。其中最主要的两个语法检查器是：基于链语法的LinkGrammar和基于错误规则（实例）的LanguageTool，我们当前的语法检查主要就是采用这两种思想的结合。另外，接下来可能要做一下三元语法，最近也抽空在研究语言模型，目前不知道效果，此处就先不介绍了。

## 基于链语法的语法检查（LG）

LG的官网<http://www.link.cs.cmu.edu/link/>。

LG是以单词为核心进行句子分析，这使得其与传统的自然语言处理句法分析相比，更易于对其进行扩展，而且其对句子局部的分析能力要远远强于传统的句法分析。在之前所作的判断句子正误的实验中，其实际效果是最好，但是缺点也很明显，就是不能有效的指出错误所在的位置（指出的位置不准确），更不能给出详细信息及修改建议。我们目前对LG的主要应用在于：（1）用其判断句子的正误，用来作为对作文评分的一方面因素，但不用其提供反馈；（2）鉴于其对句子局部的分析能力较强，我们用链接信息来对特殊句式进行判断（例如句子复杂性等）。

我们在此检查器上所作的工作主要包括以下几个方面：

1. 分析源代码，研究其工作原理。源代码位置

<http://www.link.cs.cmu.edu/link/download.html>

1. 详细分析各链接所表示的意义，并将此用作分析句子的局部与整体特性（如句子复杂度及虚拟语气等）。链接介绍<http://www.link.cs.cmu.edu/link/dict/index.html>
2. 重构代码，提高效率。（目前由上海外教社在负责进行LG部分代码的重构）

## 基于错误规则（实例）的语法检查（具体的介绍见论文）

基于错误实例的语法主要借鉴于LanguageTool，官网<http://www.languagetool.org/>。

原版的语法检查检查器效果并不是太好，在之前所做的实验中，基本上每篇作文能平均检查出一个错误，识别错误率相当低，但是它的好处是能准确的指出错误位置，给出错误信息，更易于扩展且不受句子中其他错误的影响。

对此种语法检查所作工作包括：

1. 重新设计规则文件的表示方式及检查器的核心类，LT的源代码位置

<http://sourceforge.net/scm/?type=svn&group_id=110216>

1. 扩展规则文件，目前所用的规则主要通过人工分析并加工整理完成，主要的来源包括《英语语法常见问题详解》、《薄冰英语常见共性问题解答》、《大学生英语写作指南》三本参考书及之前收集的部分科大学生语料。目前所整理的错误规则大约有300个（上次会后大约又整理了一百个左右）。
2. 提出二分匹配策略改进效率，之前做了模拟匹配效果还不错，待规则集整理完成后再进行具体实验。

## 总结及近期工作

目前基于链语法的工作主要集中在上外的重构代码部分上，我的工作主要集中在实例语法检查及接三元语法上。

关于实例语法检查：目前还有一点规则没有处理完成，预计会在三四天内完成。另外，语法检查的测试包括两部分：有效性测试和误判性测试，其中误判性测试的语料已基本准备完毕，有效性测试的语料也会在近三四天内完成。完成这些后接下来的应该就是逐个规则进行详细的测试及改进，估计约要花费十天到两周的时间。另外，尝试写一点脚本看能否半自动地获得一些有效规则。

对于三元语法：目前同时在学习语言模型，争取在两周的时间内拿出一个具体的方案。