数据库、信息处理

日语文本语义接受度评价研究

杜家利，于屏方

鲁东大学 外国语学院 汉语言文学院，山东 烟台 264025

摘要　　基于日语料库的粘着语文本语义接受度（SAS）研究分三步展开。首先提取『ゆきぐに』为分析文本，以等距离系统随机抽样方法取得6对比组。然后在屈折语SAS研究基础上提出适用于粘着语文本的词长定义，即百词所含5音拍及以上词数为超常用词量。最后得出结论：抽取间距由大变小引发抽取率（SR）由小变大的曲线变化；依次攀升的SR与围绕均值波动的SAS组图证明两者的非关联性，以实例验证了屈折语SAS评价公式对粘着语文本研究的可适用性。

关键词　　粘着语　信息检索　语料库　语义接受度　抽取率

分类号　TP311

Research on evaluation of semantic accessibility scale in Japanese text

DU Jia-li，YU Ping-fang

School of Chinese Language and Literature，School of Foreign Languages，Ludong University，Yantai，Shandong 264025，China

Abstract

The study on agglutinative-language-involved Semantic Accessibility Scale（SAS）based on Japanese corpus comprises three steps.Firstly，『ゆきぐに』 is extracted from corpus and divided into six groups for comparison by the systematic random sampling skill in which different equidistant extraction is included.Secondly，the definition of word height in presently-verified SAS formula reflecting inflecting language domain is adapted for agglutinative language domain.The word beyond five music beats is called the unpopular one，and the number of this kind of word every 100 words is considered word height.Finally，a conclusion is drawn that decreasing extracted-space results in increasing Sampling Ratio（SR），and that the non-relevance between SR and SAS is verified by the schema in which the contrast between increasing SR and the mean-fluctuated SAS is involved.In short，the evaluation of SAS in inflecting language text can be applicable in other fields，including agglutinative language text.

Key words　agglutinative language　information processing　corpus　Semantic Accessibility Scale（SAS）　Sampling Ratio（SR）

通讯作者　杜家利 dujiali68@yahoo.cn

扩展功能

本文信息

- Supporting info
- PDF(429KB)
- [HTML全文](0KB)
- 参考文献

服务与反馈

- 把本文推荐给朋友
- 加入我的书架
- 加入引用管理器
- 复制索引
- Email Alert
- 文章反馈
- 浏览反馈信息

相关信息

- 本刊中 包含"粘着语"的相关文章

本文作者相关文章

- 杜家利
- 于屏方