

数据库、信息处理

英日语料库语义接受度对比研究

杜家利, 于屏方

鲁东大学 外国语学院 汉语言文学学院, 山东 烟台 264025

收稿日期 2009-3-9 修回日期 2009-5-4 网络版发布日期 接受日期

摘要 基于语料库的语义接受度 (SAS) 研究是在线衡量文本理解程度的可行性方法。在大规模真实文本语料的基础上, 利用赋值限域方法进行英日文本对照研究。并通过分析不同赋值区间对英日小说文本语义接受度进行解读。经过验证的语义接受度公式证明了文本理解与词汇密度 (P1, P2)、词长 (H) 和句长 (L) 相关, 即 $SAS = P2 / [P1 \times 0.4 \times (L+H)]$, 而且不同的抽取率不会引起评价值的显著差异。此公式为文学研究者借助网络对电子文本进行理解度评价提供了理论支持。

关键词

[文本](#) [语料库](#) [自然语言](#) [语义接受度](#) [信息检索](#)

分类号 [TP391](#)

Comparative study on semantic accessibility scale originating from English and Japanese corpora

DU Jia-li, YU Ping-fang

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

Abstract

The corpus-based study on Semantic Accessibility Scale (SAS) is a useful method to evaluate the acceptance of electronic texts. On the basis of large-scale natural language texts, this paper compares The Old Man and the Sea and 『ゆきぐに』 from English and Japanese corpora by means of the information retrieval and semantic assignment. A conclusion is drawn that SAS is related to vocabulary density (P1, P2), vocabulary length (H) and sentence length (L), namely $SAS = P2 / [P1 \times 0.4 \times (L+H)]$. Correspondingly, different sampling ratios will not result in fundamental difference of SAS. This study provides the theoretical support for the literary critics to analyze the acceptance of internet-based texts.

Key words [text](#) [corpus](#) [natural language](#) [semantic accessibility scale](#) [information retrieval](#)

DOI: 10.3778/j.issn.1002-8331.2009.24.043

通讯作者 杜家利 dujiali68@yahoo.cn

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(499KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“
文本” 的相关文章](#)
- ▶ [本文作者相关文章](#)

- [杜家利](#)
- [于屏方](#)