Researching Academia.edu: 
a Quantitative Corpus-Analysis of Modal Verbs and Personal Pronouns in Scientific Papers

Jessica Dheskali
1st Semester PhD Student
jessica.kuechler@s2009.tu-chemnitz.de
Introduction

Academia.edu:

• online platform for academics
• used by undergraduate students up to professors
• community that brings together people, interests, documents from the same specific area of research and knowledge
• open access- everyone can upload everything
• used to gain and track popularity in the academic community, share documents and get in touch with other academics
Methodology: General Information

Sample and Analysis

- corpus analysis with AntConc (2011)
- corpus of scientific articles from Academia.edu, mainly published (in journals)
- 4 affiliations: Faculty Member/ Department Member, Post-Doc, PhD Student, Graduate Student
- 7 different nationalities (British; German, Italian; Russian, Chinese; Ghanaian, Nigerian)
- 20 different Universities
- papers ranging from 2009-2015
- field: English Linguistics and Literature
Methodology: Corpus Compilation Problems

Text selection: platform problems
- different names for same university and department
- various departments with same or similar field
- unorganized and limited search options

Text provision: author problems
- paper written in mother tongue
- co-authors from different countries not mentioned
- paper not downloadable (Schmied 2013: 152)
- missing classification of document type
Methodology: Corpus Compilation Problems

Text processing: technical problems
- OCR programs needed for scanned papers (not all directly usable with AntConc)
- limitation in word count function of AntConc (2011)
- limitation in handling of annotated data in AntConc (2011)
- inevitable manual intervention in the process of transcription (McEnery, Ostler 2000: 411)

Corpus design issues:
- low representativeness of corpus (small sample) (Biber, Jones 2009: 1288)
- internally variation of characteristics of a text: differences in (argumentation) structure (Biber, Jones 2009: 1289)
The Notion of Free Indirect Discourse and its Use in Contemporary Journalism

by Olga Blinova

Free Indirect Discourse (FID) is traditionally considered a device limited to literary texts. This article argues that FID was widely used in journalist writing throughout the XXth century and is still popular today. The analysis focuses... more

Publication Date: Dec 2012

Research Interests: Journalism, Stylists, Discourse Stylists, Linguistic stylistics, Ernest Hemingway, and 6 more

Download (.pdf)
www.universitypublications.net/hssr/0102/350442563c1c49f6b2b9fa7297c4f5c0.pdf

Bookmark

View on universitypublications.net

The Notion of Free Indirect Discourse and its Use in Contemporary Journalism

by Olga Blinova

Free Indirect Discourse (FID) is traditionally considered a device limited to literary texts. This article argues that FID was widely used in journalist writing throughout the XXth century and is still popular today. The analysis focuses... more

Publication Date: Dec 2012

Research Interests: Journalism, Stylists, Discourse Stylists, Linguistic stylistics, Ernest Hemingway, and 6 more

Download (.pdf)
www.universitypublications.net/hssr/0102/350442563c1c49f6b2b9fa7297c4f5c0.pdf

Bookmark

View on universitypublications.net
Olga Blinova

Moscow State Institute of International Relations, English Language #3,

Address:
Moscow, Russia

62 followers | 37 following | 2,358 total views | top 5%

- Follow
- Message

PAPERS IN ENGLISH

Speech and Thought Representation in Hemingway: The Case of Free Indirect Discourse

 Bookmark | Download | 100 | More

The Notion of Free Indirect Discourse and its Use in Contemporary Journalism

 Bookmark | Download | 988 | More

PAPERS IN RUSSIAN

Речевые характеристики персонажей-иностранных в романе Э. Хемингуэя "По ком звонит колокол"
Methodology

Academia.edu-Corpus
- Total number of papers: 68
- Total number of words: 425 145

<table>
<thead>
<tr>
<th>Nationality</th>
<th>Total nr. of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ghanaian</td>
<td>41274</td>
</tr>
<tr>
<td>Nigerian</td>
<td>50772</td>
</tr>
<tr>
<td>British</td>
<td>59033</td>
</tr>
<tr>
<td>Chinese</td>
<td>51885</td>
</tr>
<tr>
<td>German</td>
<td>76290</td>
</tr>
<tr>
<td>Italian</td>
<td>84035</td>
</tr>
<tr>
<td>Russian</td>
<td>61856</td>
</tr>
</tbody>
</table>

RQ1: How is the distribution of personal pronouns among the different nationalities?

RQ2: Which modal verbs are used to convey ideas, to put forward arguments and to make claims?
## Results

<table>
<thead>
<tr>
<th></th>
<th>relative frequency per 100T words</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>may</em></td>
<td>121.6</td>
</tr>
<tr>
<td><em>might</em></td>
<td>40.5</td>
</tr>
<tr>
<td><em>should</em></td>
<td>66.3</td>
</tr>
<tr>
<td><em>must</em></td>
<td>50.6</td>
</tr>
<tr>
<td>1st person SG</td>
<td>412</td>
</tr>
<tr>
<td>1st person PL</td>
<td>387.2</td>
</tr>
<tr>
<td>2nd person SG</td>
<td>145.6</td>
</tr>
</tbody>
</table>

Table 1: Relative frequencies per 100,000 words in the Academia.edu-Corpus
### Results

<table>
<thead>
<tr>
<th></th>
<th>British</th>
<th>Chinese</th>
<th>German</th>
<th>Italian</th>
<th>Russian</th>
<th>Ghanaian</th>
<th>Nigerian</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>may</em></td>
<td>118.6</td>
<td>100.2</td>
<td>98.3</td>
<td>204.7</td>
<td>87.3</td>
<td>164.8</td>
<td>69</td>
<td>121.6</td>
</tr>
<tr>
<td><em>might</em></td>
<td>59.3</td>
<td>52</td>
<td>72.1</td>
<td>35.7</td>
<td>25.9</td>
<td>12.1</td>
<td>7.9</td>
<td>40.5</td>
</tr>
<tr>
<td><em>should</em></td>
<td>40.7</td>
<td>86.7</td>
<td>69.5</td>
<td>53.6</td>
<td>74.4</td>
<td>63</td>
<td>84.7</td>
<td>66.3</td>
</tr>
<tr>
<td><em>must</em></td>
<td>40.7</td>
<td>63.6</td>
<td>23.6</td>
<td>31</td>
<td>27.5</td>
<td>75.1</td>
<td>130</td>
<td>50.6</td>
</tr>
<tr>
<td>1st person SG</td>
<td>833.4</td>
<td>354.6</td>
<td>458.8</td>
<td>643.8</td>
<td>646.7</td>
<td>697.8</td>
<td>283.6</td>
<td>412</td>
</tr>
<tr>
<td>1st person PL</td>
<td>340.5</td>
<td>235.1</td>
<td>347.4</td>
<td>363</td>
<td>679</td>
<td>438.5</td>
<td>291.5</td>
<td>387.2</td>
</tr>
<tr>
<td>2nd person SG</td>
<td>164.3</td>
<td>104.1</td>
<td>108.8</td>
<td>126.1</td>
<td>106.7</td>
<td>317.4</td>
<td>161.5</td>
<td>148.7</td>
</tr>
</tbody>
</table>

Table 2: Relative frequencies per 100,000 words in the Academia.edu-Corpus by the different nationalities
Results- Modal Verbs

Fig. 2: Relative frequencies of modal verbs in the Academia.edu-Corpus
Results - Personal Pronouns

Fig. 2: Personal Pronouns in the Academia.edu-Corpus per 100,000 words
Discussion

(1) What are the reasons for the high number of the 1st person PL pronoun in the Russian papers?

(2) Is *we* rather used inclusively or exclusively in Russian papers? (Wales 1996: 63)

(3) What are reason for highest number of hits of *must* in the Ghanaian and Nigerian papers?

(4) How can we explain the high number of hits for *may* in Italian papers?
### Results - without multi-authored papers (Russian)

<table>
<thead>
<tr>
<th>Nationality</th>
<th>British</th>
<th>Chinese</th>
<th>German</th>
<th>Italian</th>
<th>Russian</th>
<th>Ghanian</th>
<th>Nigerian</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person PL</td>
<td>340,5</td>
<td>235,1</td>
<td>347,4</td>
<td>363</td>
<td><strong>488,2</strong></td>
<td>438,5</td>
<td>291,5</td>
<td>387,2</td>
</tr>
</tbody>
</table>

Table 2: Relative frequencies of 1st person PL pronoun per 100,000 words in the Academia.edu Corpus the different nationalities

- 13 papers
- 6 with multiple authors (21029 words)
- 7 with single authors (40827 words)
- still highest number and above average
References


Researching Academia.edu: a Quantitative Corpus-Analysis of Modal Verbs and Personal Pronouns in Scientific Papers

Jessica Dheskali
1st Semester PhD Student
jessica.kuechler@s2009.tu-chemnitz.de
## List of Universities

<table>
<thead>
<tr>
<th>University of Reading</th>
<th>Albert-Ludwigs-Universität Freiburg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Justus-Liebig-University Giessen</td>
<td>Università degli Studi di Torino</td>
</tr>
<tr>
<td>University of Pavia</td>
<td>Russian State University for the Humanities</td>
</tr>
<tr>
<td>Università Ca' Foscari Venezia</td>
<td>Saratov State University</td>
</tr>
<tr>
<td>Moscow State University</td>
<td>Shanghai Jiao Tong University</td>
</tr>
<tr>
<td>Moscow state institute of international relations</td>
<td>Hexi University</td>
</tr>
<tr>
<td>Peking University</td>
<td>University of Ibadan</td>
</tr>
<tr>
<td>Fudan University</td>
<td>SOAS University of London</td>
</tr>
<tr>
<td>University of Ghana, Accra, Legon</td>
<td>Queen Mary, University of London</td>
</tr>
<tr>
<td>University of Sussex</td>
<td>Russian Academy of Sciences</td>
</tr>
</tbody>
</table>