News Genres: the potential and problems of automatic genre classification for newspaper and television news

Kim Smeenk, Frank Harbers & Marcel Broersma

Centre for Media and Journalism Studies
University of Groningen
Journalism history

> From opinionated to fact-centred journalism
> In the Netherlands between 1960 – 1990
NEWSGAC: Advancing Media History by Transparent Automatic Genre Classification

Marcel Broersma (RUG), Jacco van Ossenbruggen (CWI), Frank Harbers (RUG), Johan Oomen (NISV), Martijn Kleppe (KB), Laura Hollink (CWI), Aysenur Bilgin (CWI), Erik Tjong Kim Sang (eScience Center) & Kim Smeenk (RUG)
NEWSGAC

Three goals:

› To map the use and development of journalistic genres between 1950 and 1990

› To create a tool that is able to automatically classify the genre of historical newspaper articles and television news items

› To create a transparent dashboard for researchers as a way to make informed choices about the different pipelines and their biases in assigning genre
Genre

> “language use in a conventionalized communicative setting in order to give expression to a specific set of communicative goals of a disciplinary or social institution, which give rise to stable structural forms” (Bhatia 2004)
Genre & machine learning

› Too much research material to manually annotate

› How do we define stable genre categories?
Challenge 1: analysing genre over time

› Genres are historically situated

› We need stable categories that are flexible enough to account for internal historical changes
Challenge 2: analysing genres across media

› TV genre studies argues that genre is medium specific

› Journalistic genres are *field*-specific
## Challenge 3: from abstract concept to concrete features

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 - News article</strong> [Nieuwsbericht]</td>
<td>Factual rendition of newsworthy events; usually short and concise. Signal: the first sentence describes the ‘pointe’ of an article (this is especially applicable for the period after WWII)</td>
</tr>
<tr>
<td><strong>2 - Report</strong> [verslag]</td>
<td>Factual, chronological rendition of clearly delimited (sometimes planned) events (e.g. sports competition, parliamentary session)</td>
</tr>
<tr>
<td><strong>3 - Background / news analysis</strong> [Achtergrond / nieuwsanalyse]</td>
<td>Closer look at newsworthy events. Are an addition to the news articles, but are distinguished from them because analysis, interpretation and explication have a more important place than the facts themselves. Signal: the first sentence usually doesn’t describe the essentials of an article but most likely explains the reason.</td>
</tr>
<tr>
<td><strong>4 - Interview</strong> [interview]</td>
<td>Article that renders the primary meeting where the author visits the interviewee. The piece is the representation of a conversation, which can be told by the form.</td>
</tr>
<tr>
<td><strong>5 - Reportage/Feature</strong> [reportage]</td>
<td>Article that apart from the representation of an event also represents the corresponding experience. Usually using imagery and renderings of the atmosphere. Clues are: metaphors/imagery, presence of a storyteller (I-figure), tension build-up, attention to detail and illustrating by using quotations.</td>
</tr>
<tr>
<td><strong>6 - Opinion article</strong> [opiniestuk]</td>
<td>Author gives his opinion about an topical and newsworthy theme and gives arguments for it.</td>
</tr>
<tr>
<td><strong>7 - Review</strong> [recensie]</td>
<td>Reasoned opinion on a cultural product (e.g. concert, book, exhibition). It is valued or evaluated.</td>
</tr>
<tr>
<td><strong>8 - Column</strong> [column]</td>
<td>Regularly appearing, form-retaining, signed piece where the personal impression is most important. It could be an opinion piece or a subjective description of one or multiple events, but it could also be an elaborate recipe.</td>
</tr>
</tbody>
</table>
**Challenge 3: from abstract concept to concrete features**

| 5 - Reportage/Feature [reportage] | Article that apart from the representation of an event also represents the corresponding experience. Usually using imagery and renderings of the atmosphere. Clues are: metaphors/imagery, presence of a storyteller (I-figure), tension build-up, attention to detail and illustrating by using quotations. |
Challenge 3: from abstract concept to concrete features
Challenge 3: from abstract concept to concrete features
Conclusion

› Stable genre categories have to be defined:
  - historically
  - in relation to the relevant field

› Stabilizing genre categories enables us to uncover historical patterns