Literature Reloaded: using databases to explore literary trends

Lucija Dačić
lucija.dacic@bristol.ac.uk
Existing research on genre perception relies on stereotypes and ‘common knowledge’

Reports on genre popularity rely almost exclusively on sales data

**No clear definitions**: genres are fuzzy sets, easy to recognise at the centre but vague near the edge

**No standard labelling**: publishers use broad categories
“Works presenting modes of being that contrast with their audiences’ understanding of reality, [with] key emphasis on speculative representation of what would happen had the actual chain of events or the matrix of reality-conditions been replaced with other conditions.”

- Gill 2013
Cambridge English Corpus

- A multi-billion word collection of written, spoken and learner texts compiled by Cambridge University Press.

- Primarily used in-house to inform English Language Teaching department and publications.

- Access is restricted to researchers working for Cambridge University Press and Cambridge English Language Assessment.
Tools:

- **Word sketch difference** (how words differ in their context and behaviour)
- ‘**Thesaurus**’ (words used in similar context)
- **Word sketch** (important co-occurring words)
Cambridge English Corpus

Problems:

- Number of entries falls sharply after 2008
- Small sample size
- **Nielsen BookScan**

---

- A commercial service aimed at publishing houses that gathers weekly sales figures from various UK booksellers.
- Their statistics capture over 90% of British print book market.
- Access is restricted to institutions who are Nielsen BookScan subscribers.
Problems:

- Only rudimentary ebook and self-published market tracking
- Tracking by ISBN - bestseller charts do not always reflect actual bestselling titles
- Broad genre categories
Goodreads

- The world’s largest site for readers and book recommendations
- Over 30 million members
- Over 900 million books
- Users can categorise books with no pre-set labels or category names
Goodreads

- Tracking popularity without having to rely exclusively on book sales data

- Tracking perception of genre through user-created labels

- Combined with Nielsen data: an insight into popularity of genres and discrepancies in use of genre labels
Thank you!