

Type of Contribution: Work-in-Progress Report

Think it. Say it. Reported.

A century of reporting verbs in British fiction (1900-2019)

https://www.usit.uio.no/om/organisasjon/itf/ds/ansatte/jarlee/person/ICAME41_reporting_verbs.pdf

Ebeling, Jarle, University of Oslo, jarle.ebeling@usit.uio.no

Abstract

This work-in-progress report investigates reporting verbs in 20th and 21st century British fiction. The investigation is limited to reporting verbs in direct speech (DS) and thought (DT), e.g. *cried* in (1) and *thought* in (2), respectively.

1) "Give us another call!" he cried.

2) "Why not?" he thought.

In his book on Dickens and the Suspended Quotation, Lambert (1981) observes that the use of *cry* as a reporting verb has been, and still is, dropping in use. Based on his observation I wanted to find out whether this decrease affects only some reporting verbs, and whether some verbs become more frequent as others decrease in use.

And, if we can detect a decrease in the use of certain reporting verbs, is this tendency equally strong regardless of who is reported speaking. This last question latches on to something Underwood (2019, p.123) and Underwood et al. (2018) address, namely whether we can observe a "growing blurriness of gender boundaries" in fiction?

With these research questions as my point of departure, reporting verb and speaker (*he* or *she*) were extracted from passages of dialogue in the the Corpus of British Fiction (CBF). The CBF contains 882 novels of general fiction, by 411 different authors, and approx. 70 million words. It is tagged with the CLAWS part-of-speech tagger (Garside & Smith 1997) and lemmatised with TreeTagger (Schmid 1994). Table 1 gives an overview of number of texts/words per decade.

Table 1. The number of texts and words per decade

| Decade of publication | Number of texts | Number of words |
|-----------------------|-----------------|-------------------|
| 1900-1909 | 80 | 7,444,527 |
| 1910-1919 | 93 | 7,465,198 |
| 1920-1929 | 109 | 9,386,574 |
| 1930-1939 | 93 | 7,817,215 |
| 1940-1949 | 50 | 4,436,032 |
| 1950-1959 | 55 | 3,781,486 |
| 1960-1969 | 59 | 4,500,186 |
| 1970-1979 | 54 | 4,060,160 |
| 1980-1989 | 83 | 4,940,756 |
| 1990-1999 | 88 | 5,479,874 |
| 2000-2009 | 54 | 5,319,611 |
| 2010-2019 | 64 | 6,363,988 |
| Total | 882 | 70,995,607 |

494,000 reporting verbs were automatically extracted from the corpus. The speaker is *he* in 26% of the cases, *she* in 19% and other/unknown in the remaining cases. The unknown category is typically made up of other pronouns and proper nouns.

Figure 1 traces the amount of reporting verbs in the 120-year period covered by the CBF. The numbers have been normalised by dividing the frequency of verbs per decade by the total number of both direct and free direct speech and thought in that decade.

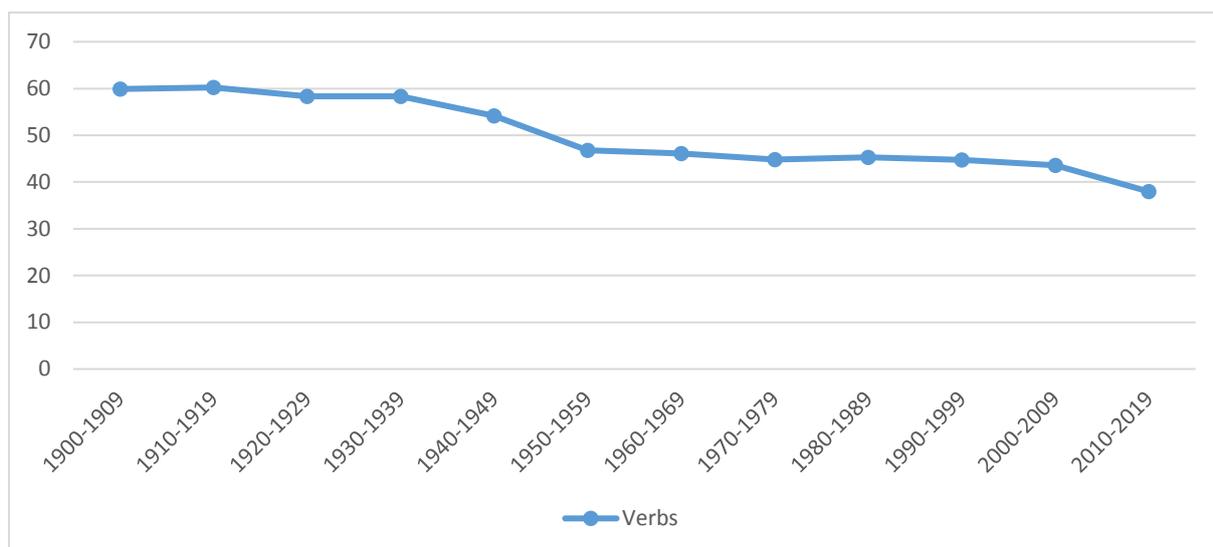


Figure 1. Reporting verbs per decade

Manual scrutiny of 5,000 randomly selected instances of both direct and free direct speech and thought showed that a little more than half of all occurrences (57%) are free direct speech and thought, as in *"Oh, for God's sake."*

The figure shows a steady decrease in the use of reporting verbs from 1900 to the present. There may be several reasons for this, e.g. the structure of the corpus or the ways in which direct speech is marked up. A manual inspection of 2,000 randomly selected positives and negatives revealed 93% precision and 74% recall.

As has been noted by several scholars, e.g. de Hahn (1996), *say* is the reporting verb par excellence, and, indeed, it accounts for more than 60% of all occurrences of reporting verbs in the CBF (300,000 of the 494,000). The second-most frequent verb is *ask* which accounts for 5.5% of the occurrences. *Whisper* at rank ten accounts for 0.9% and *laugh* at rank thirty 0.3%.

Reporting verbs sorted by frequency: *say* (60.7%), *ask* (5.5%), *reply*, *cry*, *answer*, *tell*, *add*, *exclaim*, *murmur*, *whisper* (0.9%), *remark*, *go on*, *continue*, *think*, *repeat*, *agree*, *begin*, *shout*, *explain*, *mutter*, *demand*, *enquire*, *call*, *suggest*, *observe*, *declare*, *admit*, *protest*, *announce*, *laugh* (0.3%)

The infrequency of verbs ranked below the top ten in the later decades means that we get few instances per decade to draw our conclusions on, and this number becomes even lower when we only look at the ones where *he* or *she* is the speaker.

To investigate Lambert's claim about the decreasing use of *cry* as a reporting verb, the data were split into three periods, 1900-1939, 1940-1979 and 1980-2019, and sorted by raw frequency.

1900-1939

*say, ask, reply, **cry**, answer, exclaim, remark, murmur, think, whisper, add, continue, repeat, go on, enquire, begin, agree, demand, explain, mutter, tell, suggest, observe, declare, laugh, shout, admit, call, retort, return*

1940-1979

*say, ask, tell, reply, add, answer, **cry**, exclaim, agree, go on, murmur, think, explain, shout, whisper, repeat, call, begin, remark, demand, suggest, enquire, continue, mutter, announce, protest, snap, smile, admit, sigh*

1980-2019

*say, ask, tell, reply, add, whisper, shout, continue, murmur, mutter, call, go on, begin, agree, snap, explain, **cry**, repeat, demand, answer, suggest, exclaim, announce, enquire, yell, protest, admit, comment, remark, observe*

A quick perusal of the lists shows that there is remarkable stability in the use of reporting verbs in the period covered. In fact, several scholars (e.g. Kytö et al. 2006; Busse forthc.) make this point and argue that we should not disregard stability as a powerful factor in the history of language in our (eager) pursuit of change. We can notice some changes, though, not least in relation to *cry*. *Cry* has dropped in popularity, while another verb with overlapping meaning, *shout*, has gained in popularity. Is it the case that *shout* is about to oust *cry* as a way of expressing a strong emotion in British fiction?

The stability is also apparent when ranking the verbs according to speaker, *she* or *he*, for the whole period (1900-2019).

Female speaker (*she*)

*say, ask, reply, cry, answer, tell, murmur, add, whisper, think, exclaim, agree, repeat, remark, go on, begin, continue, explain, shout, enquire, mutter, call, suggest, demand, observe, **protest**, admit, snap, declare, **smile***

Male speaker (*he*)

*say, ask, reply, cry, answer, tell, exclaim, add, murmur, remark, go on, whisper, continue, repeat, agree, begin, think, shout, mutter, explain, demand, enquire, call, suggest, **observe**, declare, **announce**, admit, laugh, snap*

If we compare the two lists, only *protest* and *smile*, which occur among the top 30 verbs with *she*, do not occur among the top 30 with *he*. Similarly, two verbs used by *he* do not make it to the top 30 list used by *she*: *observe* and *announce*. More interesting perhaps is the rank of *think*. If we take these two lists at face value, *she* seems to be thinking more than *he*, or, more accurately, authors depict women through direct thought more than they do men.

Regarding the potentially "growing blurriness of gender boundaries", this is difficult to establish based on the CBF due to the drop in overall frequency of verbs other than *say*. Figure 2, for instance, showing the development of the use of *think*, seems to indicate this blurriness of gender boundaries, but as there are few occurrences of *he/she thought* in the later decades, more data is needed to establish blurriness of gender with more certainty.

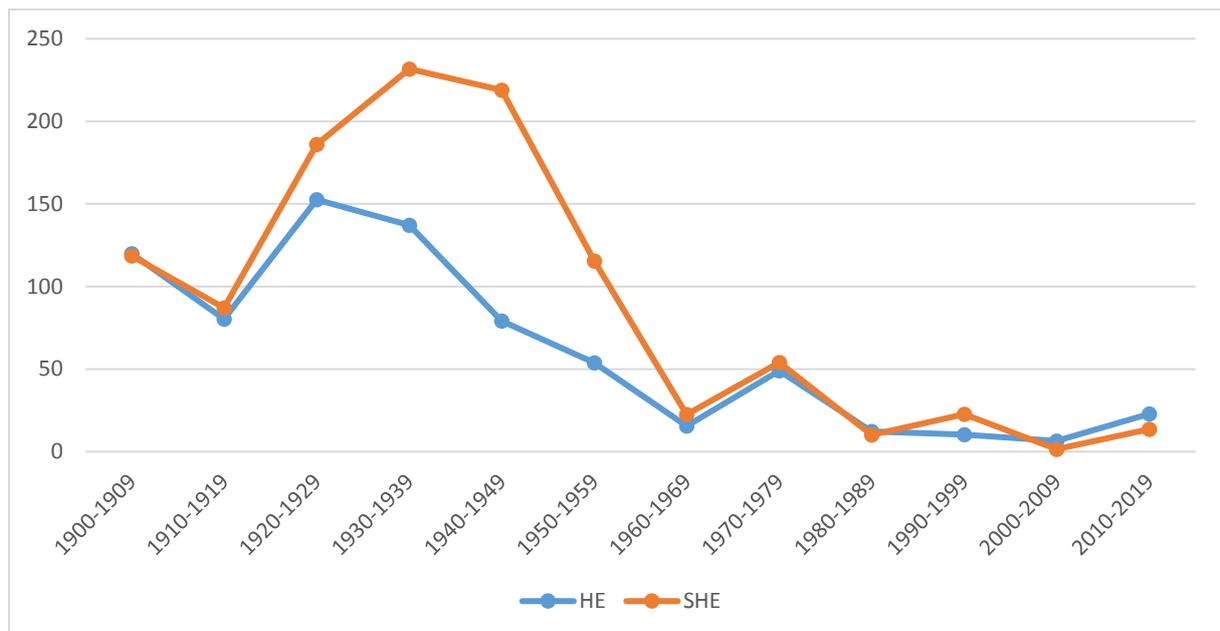


Figure 2. Normalised frequency of *he* / *she* THINK

What Figure 2 does show, however, is that the use of direct thought was more often a characteristic of females than males between 1920-1960, but seemingly not before or after that period.

- To sum up:., direct speech and thought seem to be decreasing in use, perhaps at the expense of free direct speech and thought;
- *say* is more frequently used as a reporting verb than all the other verbs put together;
- there is great stability in the use of reporting verbs in the period covered, and also when it comes to who is speaking, i.e. either *he* or *she*;
- the sharp drop in use of individual verbs overall, e.g. *think*, makes it difficult to make definite claims about any blurriness of gender boundaries, although there is some evidence that the gap between the genders is narrowing when we move closer to the present.

References

Busse, B. (Forthcoming). *Speech, Writing, and Thought Presentation in 19th-century Narrative Fiction: A Corpus-assisted Approach*. Manuscript. Oxford University Press.

- de Haan, P. (1996). More on the language of dialogue in fiction. *ICAME Journal* 20, 23–40.
- Garside, R. & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus Annotation: Linguistic Information from Computer Text Corpora*. (pp. 102–121). Longman.
- Kytö, M., Rydén M., & Smitterberg. E. (2006). Introduction: Exploring nineteenth-century English – past and present perspectives. In M. Kytö, M. Rydén, & E. Smitterberg (Eds.), *Nineteenth-Century English: Stability and Change*. (pp. 1–16). Cambridge University Press.
- Lambert, M. (1981). *Dickens and the Suspended Quotation*. Yale University Press.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of the International Conference on New Methods in Language Processing*. Manchester, UK.
<https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- Underwood, T. (2019). *Distant Horizons. Digital Evidence and Literary Change*. The University of Chicago Press.
- Underwood, T., Bamman, D., & Lee, S. (2018). The Transformation of Gender in English-Language Fiction. *Cultural Analytics* Feb. 13.
<https://DOI: 10.22148/16.019>