

# Perceived Text Quality and Readability in Extractive and Abstractive Summaries

Julius Monsen, Evelina Rennes

## Abstract

We present results from a study investigating how users perceive text quality and readability in extractive and abstractive summaries. We trained two summarisation models on Swedish news data and used these to produce summaries of articles. With the produced summaries, we conducted an online survey in which the extractive summaries were compared to the abstractive summaries in terms of fluency, adequacy and simplicity. We found statistically significant differences in perceived fluency and adequacy between abstractive and extractive summaries but no statistically significant difference in simplicity. Extractive summaries were preferred in most cases, possibly due to the types of errors the summaries tend to have.

## Data

The data used for training the models were news articles published in *Dagens Nyheter* (DN). Associated with each article was a preamble, here used as the summary.

**349,935** article-summary pairs were used for training and testing the models and producing summaries for the survey.

## Models

The **abstractive** model was trained utilising a pre-trained Swedish BERT model to warm-start an encoder-decoder model.

The **extractive** model was trained using the same pretrained Swedish BERT model as a base and the *TransformerSum* framework.

## Survey

15 articles with their respective abstractive and extractive summaries were evaluated in a web survey.

Texts were rated regarding **fluency** (a), **adequacy** (b–e), and **simplicity** (f):

- The summary contains grammatically correct sentences
- The meaning of the summary conforms to the meaning of the original text
- All the important information of the original text is contained in the summary
- The summary contains superfluous information
- The summary contains words that do not fit the context
- The summary is easy to understand

Each question was assessed using a 5-point Likert scale ranging from *Strongly Disagree* (1) to *Strongly Agree* (5).

Finally, the participants were asked which of the summaries they found to be the best and to give a short explanation of their choice.

## Results

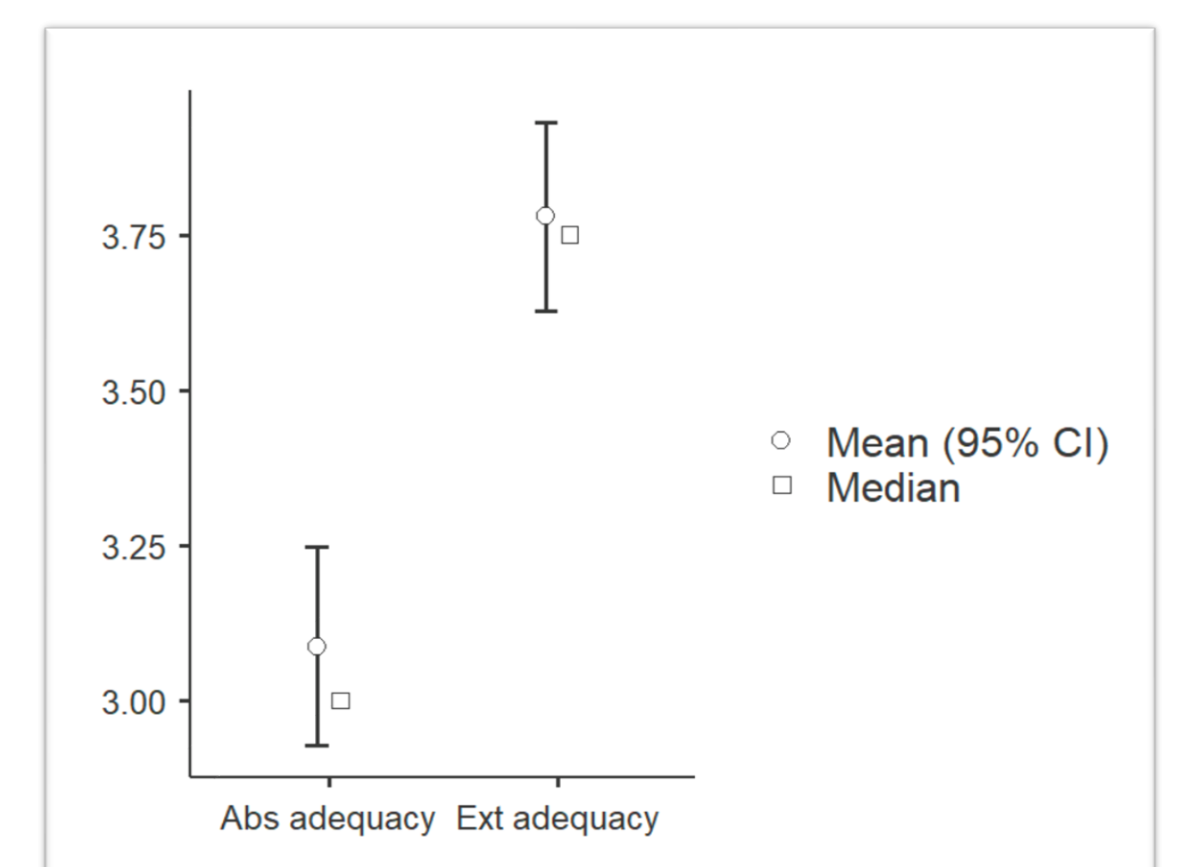
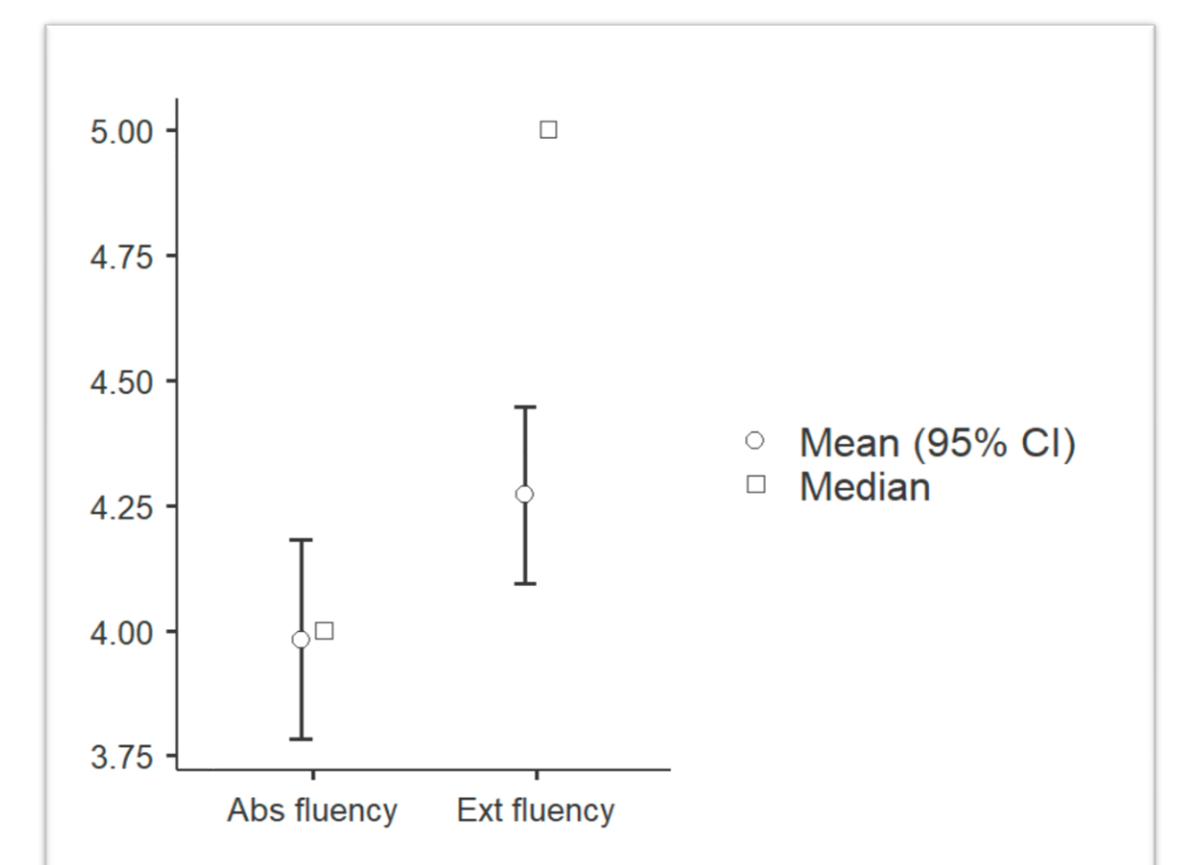
- Extractive summaries were perceived as **more fluent**.
- Extractive summaries were perceived as **more adequate**.
- No statistically significant difference was found regarding simplicity.
- Participants generally considered extractive summaries to be the best.

Question	Ext M	Abs M
(a)	4.27 (0.953)	3.98 (1.070)
(b)	3.72 (0.962)	2.51 (1.242)
(c)	3.15 (1.169)	2.40 (1.154)
(d)	3.98 (1.144)	3.46 (1.242)
(e)	4.26 (1.059)	3.98 (1.191)
(b)–(e)	3.78 (0.817)	3.09 (0.862)
(f)	3.55 (1.241)	3.41 (1.232)

Question	W(111)	Cohen's d
(a)	577*	0.271
(b)	465**	0.751
(c)	841**	0.506
(d)	582**	0.359
(e)	523*	0.209
(b)–(e)	900**	0.645
(f)	1037***	0.094

\*  $p < 0.05$ , \*\*  $p < 0.001$  \*\*\*ns

- The most common reason for preferring the extractive summary was that it was more in line with the facts presented in the article than the abstractive summary
- The main reason for choosing the abstractive summary as the best one was that the extractive summary was poorly structured and hard to follow or that it lacked some essential information.



## Conclusion

We want to emphasise the importance of the human-centred perspective on summarisation. Users' preferences are valuable to help developing summarisation systems and to make them useful in practice.

- Extractive summaries seemed to be more in line with the preferences of the participants.
- The findings highlight the need for further development of abstractive methods as they still have certain limitations.