



INTRODUCTION

Previous work on emotion-aware text simplification showed that LLMs often preserve broad emotion labels after simplification (Bezobrazova, Sokova and Orăsan, 2026). However, emotion preservation does not guarantee that the full meaning is preserved. The present study therefore focuses on meaning preservation, including core message, tone, social-media cues, and unsupported additions.

- Accessibility and plain-language guidelines exist for public services and local government.
- There is limited guidance on simplifying social media content.
- Social media adds extra challenges, such as emojis, hashtags, informal grammar, and reliance on context.



RESULTS ON MEANING PRESERVATION

Manual annotation of 50 original-simplified pairs	Emotion same	Emotion changed
Meaning preserved	17	7
Meaning not preserved	6	7
Maybe	10	3

Original	Simplified	Meaning preserved?	The issue
"sLOVEnian Power! ;)"	"Slovenian Power! 😊"	Maybe	Wordplay with "LOVE" is lost
"Yeah I regret ever moving to Leduc. It's one of the most backwoods, proud-to-be-ignorant communities I've ever seen."	"I wish I never moved to Leduc. It's a place where people are proud to be ignorant."	No	Sarcasm became literal
"Fuck my life"	"I'm really upset right now."	Maybe	Profanity and intensity softened
"Great deal! It will be my loss if I decline."	"This is a great deal! I will miss out if I say no."	Yes	"Loss" is correctly rewritten as "miss out."

Manual annotation of 50 original-simplified pairs showed that meaning was preserved in 24 cases, changed in 13 cases, and uncertain in 13 cases. This suggests that LLM simplification can improve readability, but it does not always preserve the full meaning of user-generated posts.

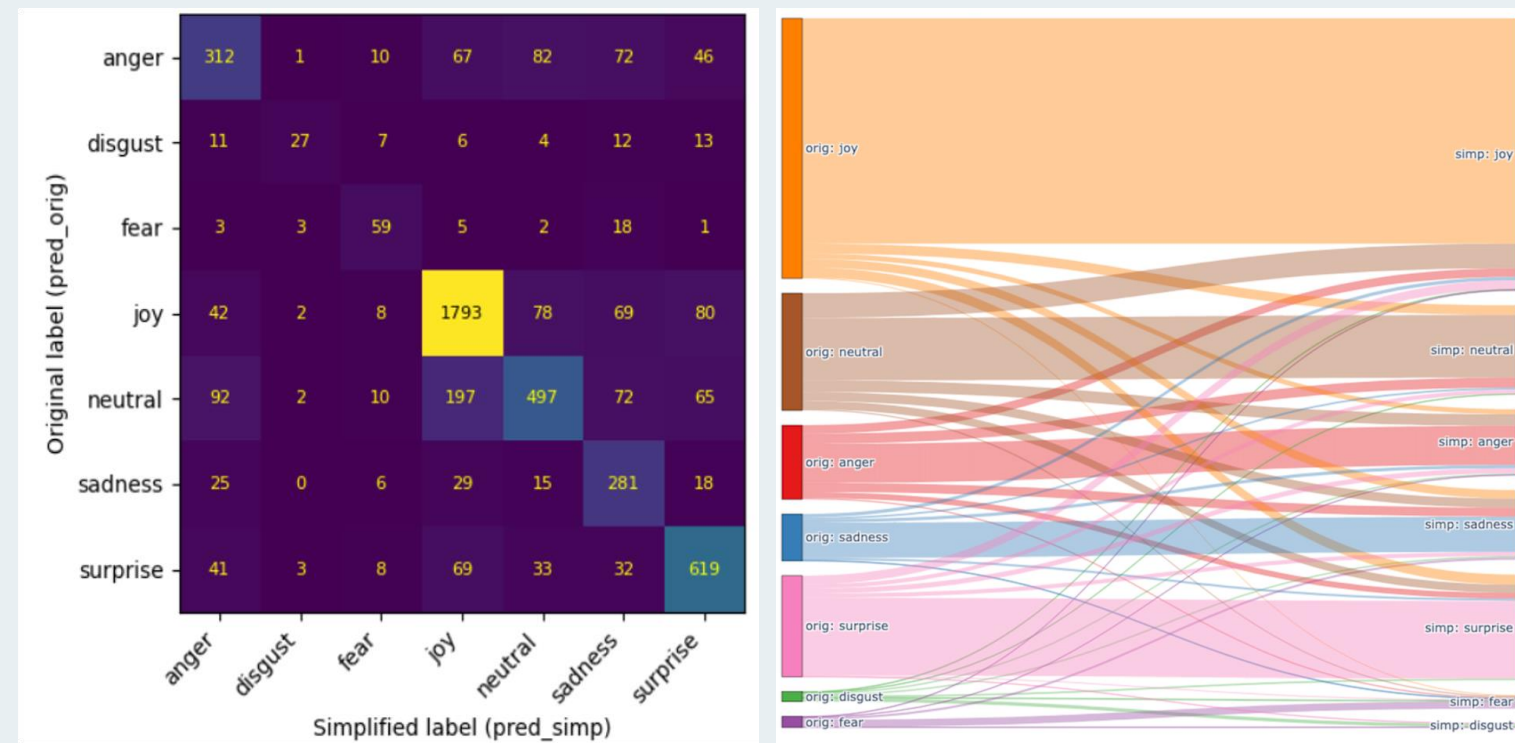
- 1. Can the original post be understood?**
If no → **MAYBE**: context unclear
- 2. Is the core meaning preserved?**
Check: what happened, who/what is discussed, main claim
If no → **NO**: core meaning changed
- 3. Are social-media cues preserved or clearly rewritten?**
Check: abbreviations, hashtags, emojis, slang, profanity, named references
If no/unsure → **MAYBE**: cue meaning unclear
- 4. Has the tone, stance, or evaluation changed?**
Examples: tone is softened, strengthened, or made more positive/negative
If yes → **NO**: stance/tone shift changes meaning
- 5. Has unsupported meaning been added?**
Examples: extra facts, advice, judgement, unsupported explanation
If yes → **NO**: unsupported interpretation added
If no → **YES**: meaning preserved

DATASET

- GoEmotions - English Reddit comments,
- Ekman-style mapping with basic emotions.
- 58,000 comments, it is curated to minimise toxicity and bias.



PREVIOUS WORK ON EMOTION PRESERVATION



Agreement between emotion predictions on original vs. simplified posts. **Left:** confusion matrix of predicted labels (original vs. simplified). **Right:** Alluvial diagram showing how predictions shift between labels.

Metric	Original	Simplified
FleschReadingEase	99.47	105.59
Kincaid	1.76	0.41
ARI	2.56	0.93
DaleChallIndex	4.07	3.37

Average readability scores for original vs. simplified posts. Automatic Evaluation Metrics cannot be applied.

Emoji & Hashtag Use (Original → Simplified)

- 😊 **Emojis:** 164 in 90 posts (<2%) → 696 in 632 posts (~13%)
- # **Hashtags:** 12 in 11 posts (<0.2%) → 706 in 692 posts (~14%)

Prompt design

Simplify the posts so that people with learning disabilities can easily understand it. Keep the same meaning and facts. Preserve the same emotion. Do not soften or exaggerate the emotion. Make the feelings clear and simple. Do not add new facts or advice. Do not judge the person. Use common words and active voice

Keep emojis only if they add meaning, and also name the feeling in words. Use CamelCase for hashtags. For example, instead of #learningdisabilityweek, write #LearningDisabilityWeek.

Alternative prompts

- Too many "I feel X" sentences
- New emojis/hashtags added
- Abbreviations/idioms not explained
- Removed "name the feeling" + added "don't add feelings" → fewer emotion statements, not zero.
- Removed hashtag instruction → fewer new hashtags (emojis may remain).
- Asked to explain abbreviations/entities → sometimes helps (e.g., "lmao"), but inconsistent; memes still unclear.

References

Anastasiia Bezobrazova, Daria Sokova, and Constantin Orasan. 2026. Emotion-aware text simplification of user generated content using LLMs. In *The Proceedings for the 15th Workshop on Computational Approaches to Subjectivity, Sentiment Social Media Analysis (WASSA 2026)*.

CONCLUSION

In the preliminary sample, meaning was preserved in 24 out of 50 cases, while 26 cases were either changed or uncertain. This supports the need for human annotation and a dedicated meaning-preservation framework rather than relying only on emotion or polarity labels.

LLM simplification may improve readability but introduce new communicative cues that were not present in the original.

Polarity shift can be a warning signal for meaning loss, but polarity stability does not guarantee meaning preservation.

Meaning preservation is difficult to evaluate because many user-generated posts are context-dependent.

Prompting is promising, but more attention should be paid to prompt engineering

Consideration of more elaborate architectures