

Embedding Emotions: Measuring What Matters in Flemish Daily Stories

Understanding the emotions embedded in everyday language requires models to effectively capture the subtlety of open-ended narratives. These unstructured text narratives reveal deeper insights into how individuals spontaneously construct and label their experiences, providing a window into the emotions hidden in daily stories. Prior works have utilized Large Language Models (LLMs) to social media datasets, but little is known about how modern neural models perform on naturally occurring narratives in low-resource languages such as Belgian Dutch (Flemish). This work addresses this gap by applying BERTopic [1], a neural topic modeling technique that leverages transformer-based contextual embeddings to a unique dataset of 24,854 daily narratives collected from 102 native Dutch speakers in Belgium over 70 days through experience sampling. Participants responded four times daily to the prompt, “What is going on now or since the last prompt, and how do you feel about it?” We performed topic modeling with (dataset-specific) emotion words masked (to avoid bias), comparing BERTopic to statistical baselines (KMeans Clustering, Latent Dirichlet Allocation (LDA)). Our evaluations followed a dual framework: (1) standard quantitative coherence metrics (such as C_v , U_{Mass} , C_{NPML} , C_{UCI}) and topic diversity, and (2) qualitative human evaluation by two Dutch speakers, directly testing whether these metrics are “measuring what matters.”

Major findings: Standard automated topic coherence metrics fail to capture emotional nuances in daily narratives with the embedding-based approach revealing superior semantic quality of topics despite lower quantitative scores:

- (A) **Metric-Quality Disconnect:** Although LDA achieved higher C_v (0.54) and topic diversity (0.96) scores than BERTopic (0.34, 0.84), human evaluation revealed that its topics often featured semantically irrelevant word co-occurrences (see Table). This could reflect its bias toward “syntagmatic associations” (statistical proximity) rather than true “paradigmatic relevance” (thematic consistency) [e.g., *bus* near *studeren* (to study)]. In contrast, BERTopic successfully uncovered culturally resonant themes intrinsic to our Flemish dataset, e.g., differentiating everyday routine ('workday routine', 'planning and communication', and 'studying') from activities ('horse riding', 'travel and outdoor recreation', and 'film evenings'), and overall emotional and mental state ('headaches and migraine-related pain', 'academic stress and assignments'). Overall, while LDA and KMeans struggled with semantic fragmentation and noise, BERTopic’s embeddings captured meaningful contextual relationships [e.g., *bibliotheek* (library) ↔ *studie_tijd* (study time)], aligning with human intuition.
- (B) **Emotion-Context Relationships:** BERTopic also revealed **one-to-many mappings between an emotion word and contexts** (e.g., *anger* spanning sports, traffic, and physical pain across 37 topics; *proud* spanning accomplishments, work meetings, walking the dog across 39 topics; *sad* spanning hospital visits, sickness, close relationships across 31 topics; *amused* spanning visiting family, supporting friends, football across 36 topics), demonstrating that emotions are inherently contextual and difficult to be captured by traditional bag-of-words approaches.

Implications: These findings challenge current evaluation paradigms in topic modeling and highlight the need for human-centered metrics for affective computing. For low-resource language processing, embedding-based approaches offer promising pathways to capture cultural specificity without sacrificing semantic coherence. Our work demonstrates that everyday emotional narratives can inform more nuanced language models for mental health applications, where understanding the context-emotion relationship(s) is crucial for effective intervention design.

	Topic No	Top words describing the topic
BERTopic	7	<i>studeren</i> , bibliotheek, tevreden, aula_vriend, studie_tijd, leeszaal, bib, dezezoachtend, basically
LDA	108	les, <i>studeren</i> , bus, snappen, interessant, deadline, toekomst, af_sluiten, practicum, focus
KMeans	73	<i>studeren</i> , goed, vandaag, eten, beginnen, rest, dag, ver, moe, tijd, oké, weten, slapen, morgen, zin

Table. Topics identified by BERTopic, LDA, and KMeans related to “studying and academic life”

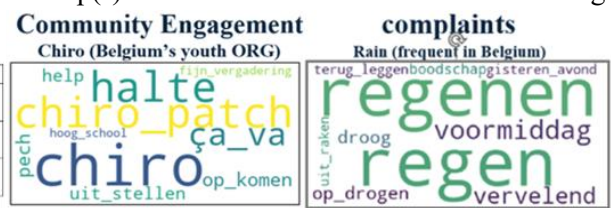


Fig. BERTopic’s word clouds – the topics were more interpretable