

Narrative Detection in Online Patient Communities

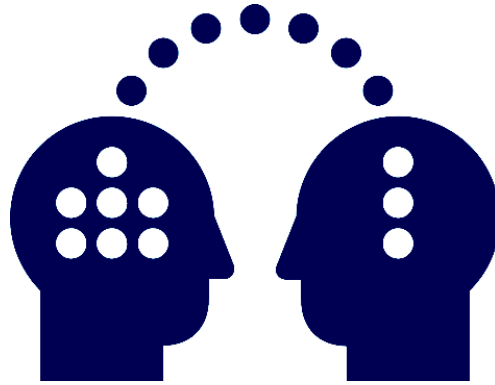
Anne Dirkson

14 April 2019 @ Text2Story



**Universiteit
Leiden**
The Netherlands

Patients share narratives on patient forums



Experiential knowledge

- Borkman (1976)¹:
'truth learned from personal experience [...]'
- Traditionally undervalued in research ^{2,3}
- Insight into Quality of Life

1. Borkman (1976). Experiential Knowledge: A New Concept for the Analysis of Self-Help Groups. *Social Service Review*, 50 (3)

2. Carter (2013). Mobilising the experiential knowledge of clinicians, patients and carers for applied health-care research. *Contemporary Social Science*, 8 (3).

3. Knotterus & Tugwell (2012). The patients' perspective is key, also in research. *Journal of Clinical Epidemiology* 65 (6)

Applications of health-related social media



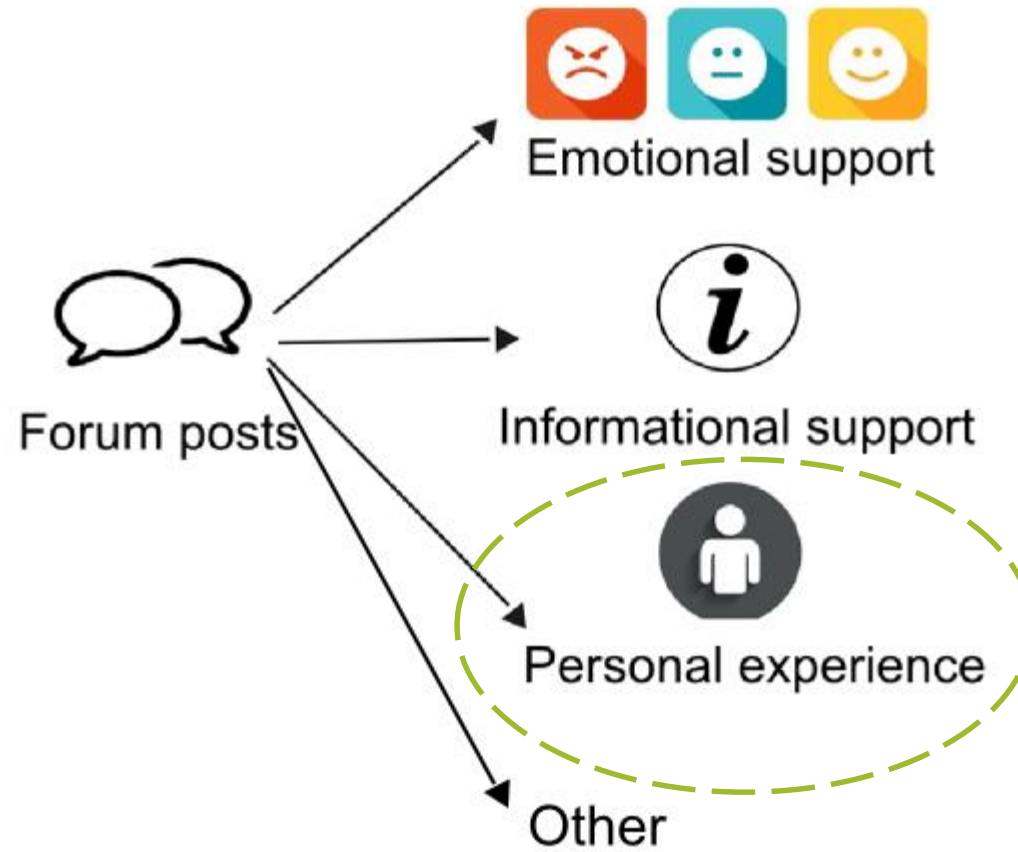
To detect drug
abuse



To track spread
of disease



To detect adverse
drug responses
(pharmacovigilance)



Automatic detection of narratives

- One study on a Dutch cancer forum ($F_1 = 0.91$)⁴
- Narratives associated with psycho-linguistic features
 - Linguistic: 1st person, 3rd person, Negations
 - Psychological: Social processes, Religion

Research questions

1. Can **document embeddings and/or psycho-linguistic features** improve the identification of patient narratives?
2. How do narratives **differ** from other types of posts?
3. How **prevalent** are narratives on a cancer patient forum and **which topics** these narratives discuss?

Data

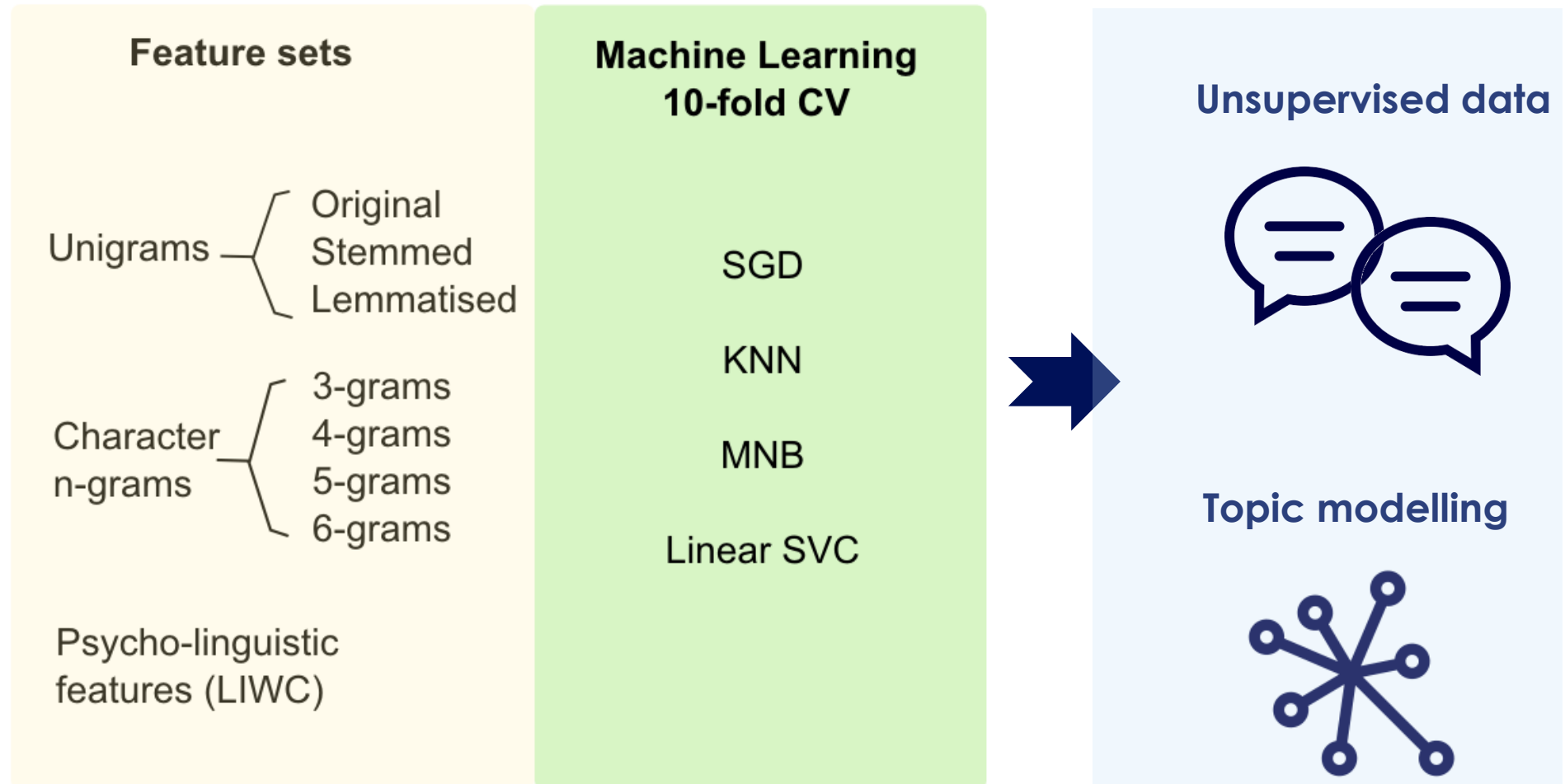
UNSUPERVISED DATA

- Cancer forum of GIST patients
- 36.722 posts

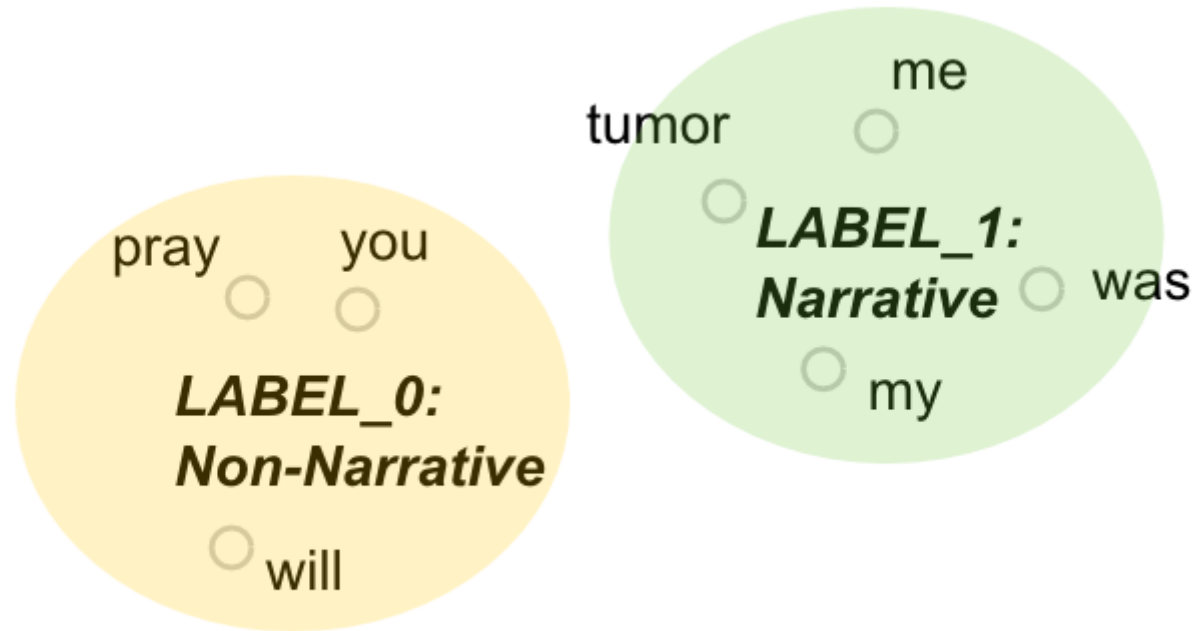
SUPERVISED DATA

- 1000 posts
- 7 annotators
- 20% overlap $\rightarrow \kappa = 0.69$
- 37.7% positive

Method



Doc2Vec Document Embeddings



Classifier

- Average vector for label
- Logistic Regression
- Dimensionality of embeddings (100 - 400)

Classifier Evaluation

Feature set		$F_1(\%)$	Recall (%)	Precision (%)
Unigram	<i>Original</i>	79.5	78.8	81.1
	<i>Stemmed</i>	81.4	79.3	84.0
	<i>Lemmatised</i>	80.8	81.0	81.3
Character n-grams	<i>3-grams</i>	81.5	84.4	79.3
	<i>4-grams</i>	81.1	82.7	84.4
	<i>5-grams</i>	79.6	78.4	81.7
	<i>6-grams</i>	79.3	79.7	79.5
LIWC		77.3	80.5	75.2
Doc2vec	<i>DBOW</i>	73.7	75.1	73.5
	<i>DM</i>	76.2	74.9	78.5
	<i>DM + DBOW</i>	77.2	80.3	74.9

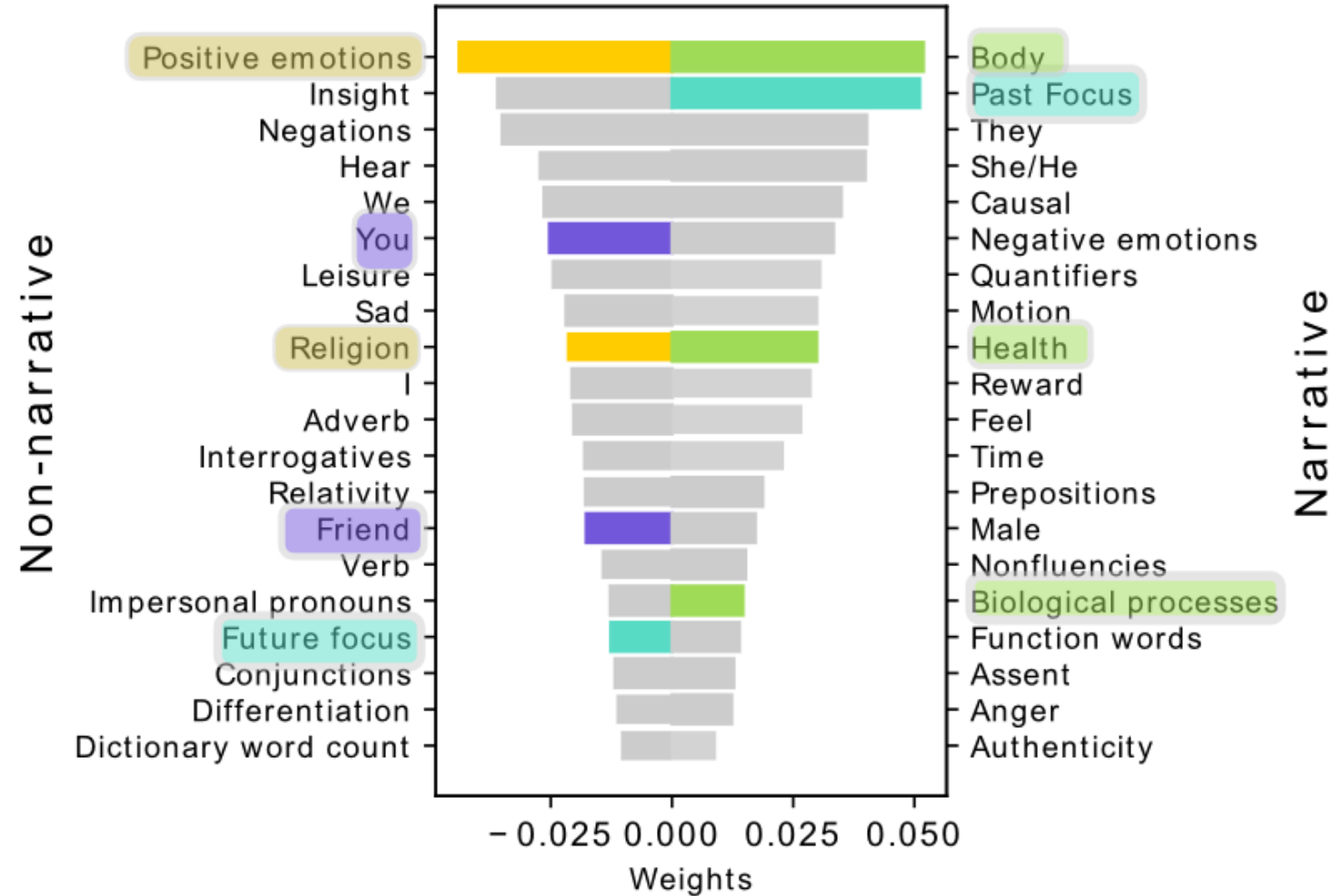
Influential features

Narratives:

- Past tense
- 1st person pronouns
- Health

Non-narrative

- Future tense
- 2nd person pronouns
- Emotional support



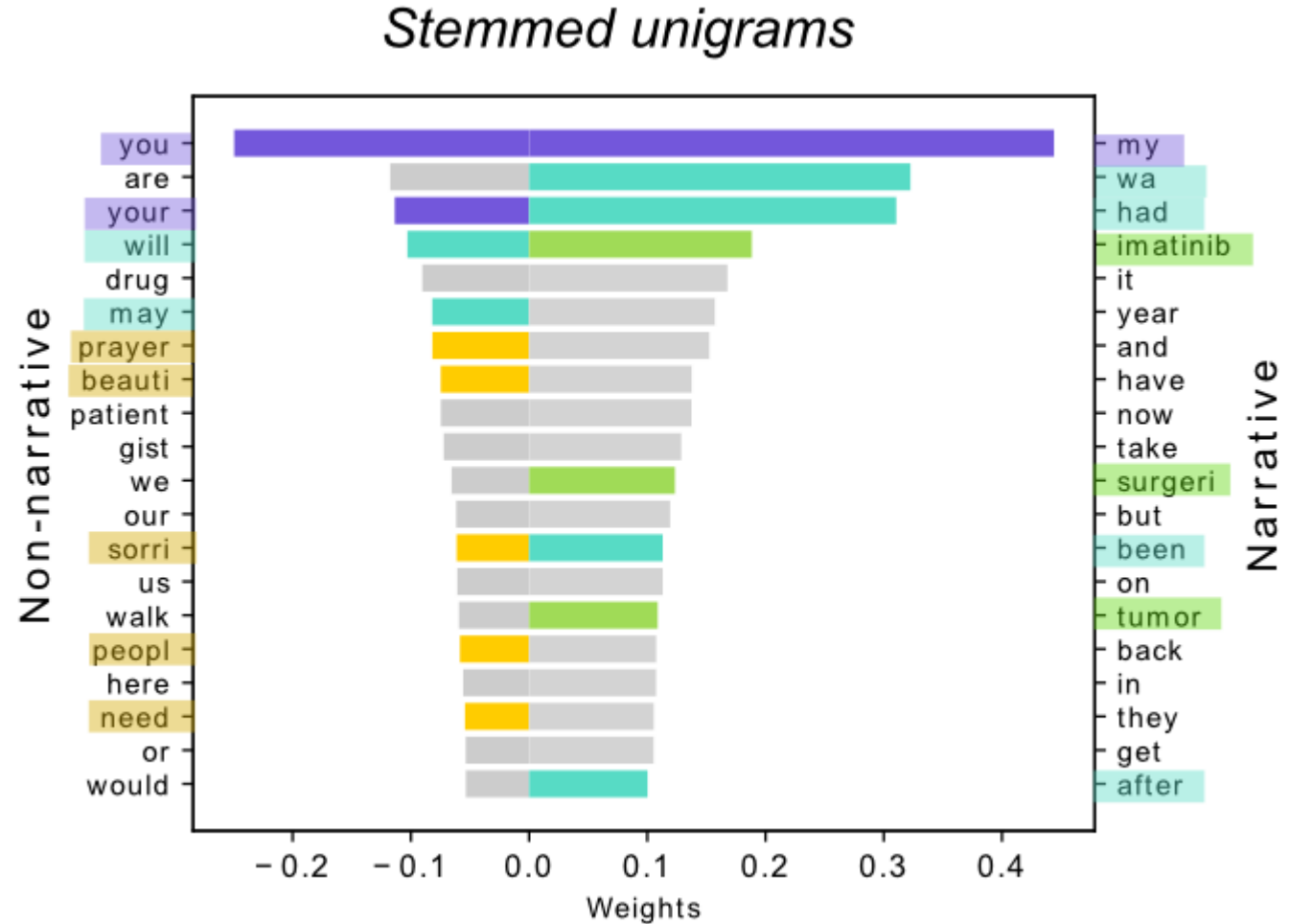
Influential features II

Narratives:

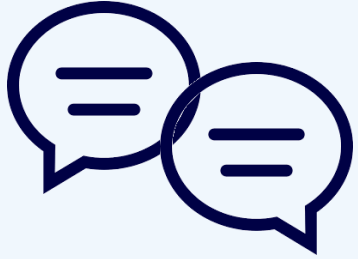
- Past tense
- 1st person pronouns
- Health

Non-narrative

- Future tense
- 2nd person pronouns
- Emotional support



Unsupervised data



37% is narratives

Topic modelling



NFM with TF-IDF weighting no stop words

TC-W2V⁶ modelling for number of topics

14 Topics

Selection of Topics

Topic label	Top 5 words	Top-ranked post
Tumor Location	Tumor stomach removed liver small	I only had one tumor on my stomach
Other Medication than First-line Treatment	Sunitinib regorafenib sorafenib imatinib working	I have this on sunitinib
Side Effects	Side effects imatinib effect different	and no side effects
Absence of Tumor Recurrence	Disease evidence still years today	No evidence of disease no evidence of disease
Dosage of Medication	Mg 400 800 imatinib 600	11 years of imatinib since 2003 at 600 mg and since November 2009 at 800 mg [...]

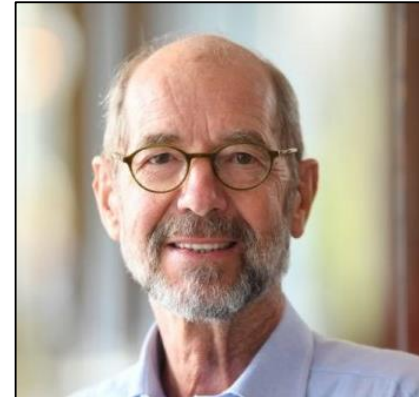
Conclusions

RQ1 • Character 3-grams are the best features for narrative classification.

RQ2 • Narratives are characterized by: 1st person pronouns, health terms and past tense

• Non-narrative text is characterized by: 2nd person pronouns, emotional support and future tense

RQ3 • 37% of a cancer forum consisted of narratives about 14 different topics





**Universiteit
Leiden**
The Netherlands



a.r.dirkson@liacs.leidenuniv.nl



github.com/AnneDirkson