



# Natural Language Processing (NLP) in Stakeholder & Consumer Insight Research to complement interpretivist research approach



**Jian Liu, Jenny Young, Tracey Phelps, Denise Conroy and Ivy Gan**



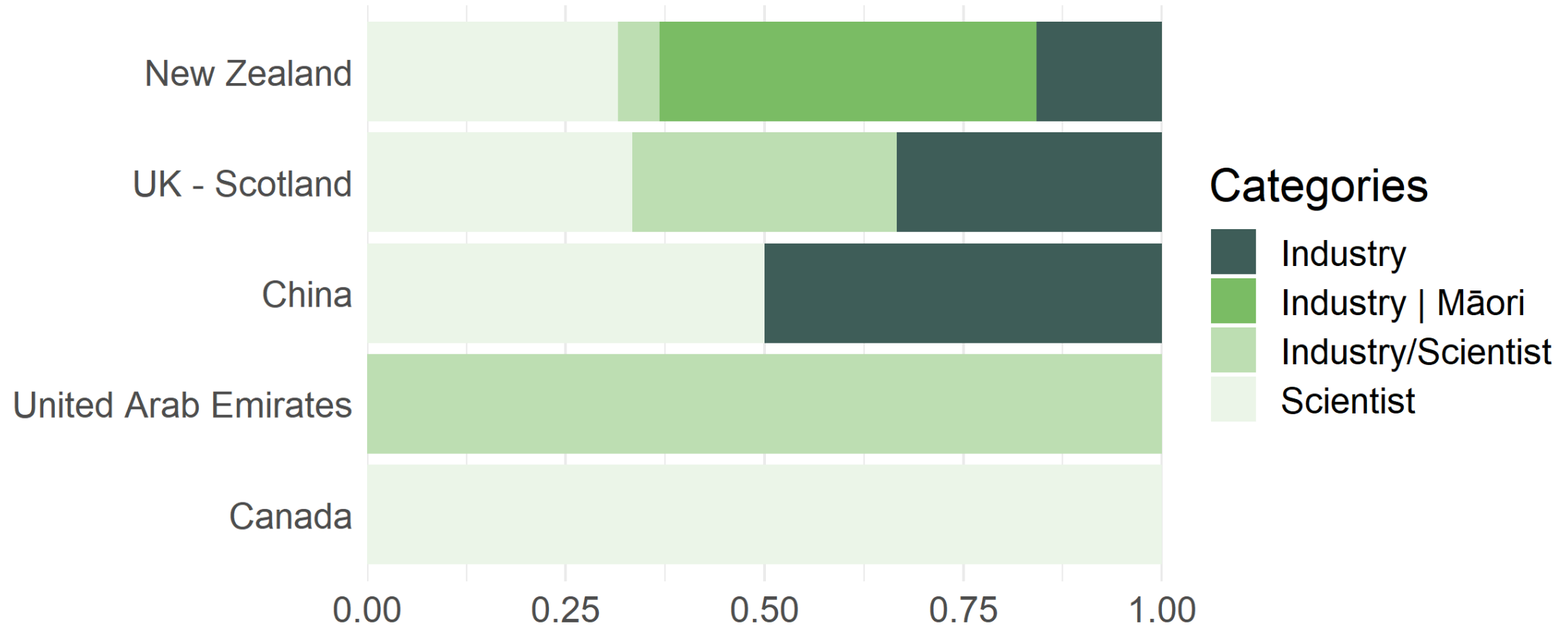
## » High level research question:

How do expert stakeholders (local and international, in both science and industry) currently understand and make sense of Controlled Environment Agriculture (CEA) as a future food technology?

## » What NLP can help to complement qualitative research?



## Sampling distribution in five countries.



Proportion

Total number of samples is 30.

# A snapshot of the raw data



<b>Speaker 1</b> <b>01:10</b>	Greetings. Could you introduce your background a little bit?
<b>Speaker 2</b> <b>02:00</b>	I graduated from XXX in XXX. My current interest .... My current role at XXX is XXX.
<b>Speaker 1</b> <b>07:30</b>	Thank you. What is your opinion on topic A?
<b>Speaker 2</b> <b>08:00</b>	Er, I think overall that I have a mixed feeling about topic A. One side is that.... Another thing is that ... I think .... I know that ... From my experience, topic A has ....
<b>Speaker 1</b> <b>15:10</b>	Really interesting insights you just mentions. I think you are most right on xxx. Could you extend on the second point a bit more?
<b>Speaker 2</b> <b>16:10</b>	Yeah, for sure. What I mean by xxx is that ..... So the situation will be complicated if we assume that the customer have some perceptions on ..... I think it is hard to predict what future will look like.
<b>Speaker 3</b> <b>18:55</b>	Totally agreed. Predicting future is very very difficult. So you suggest that xxx. Why would you think that is the case?
<b>Speaker 1</b> <b>19:30</b>	Em... a couple of points there. First, ..... Then, ..... I think the last one maybe around .....

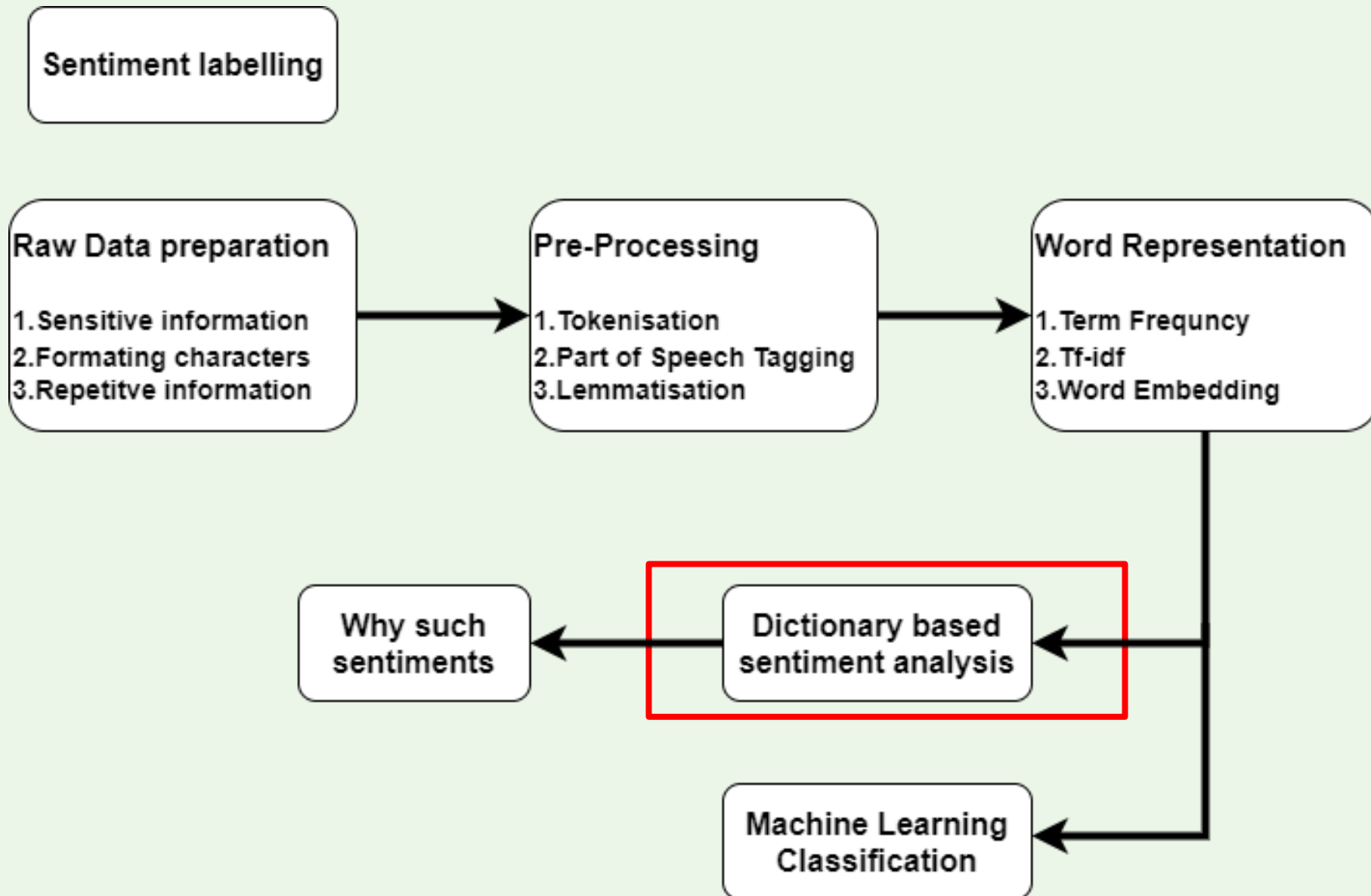
## Some challenges in the pre-processing



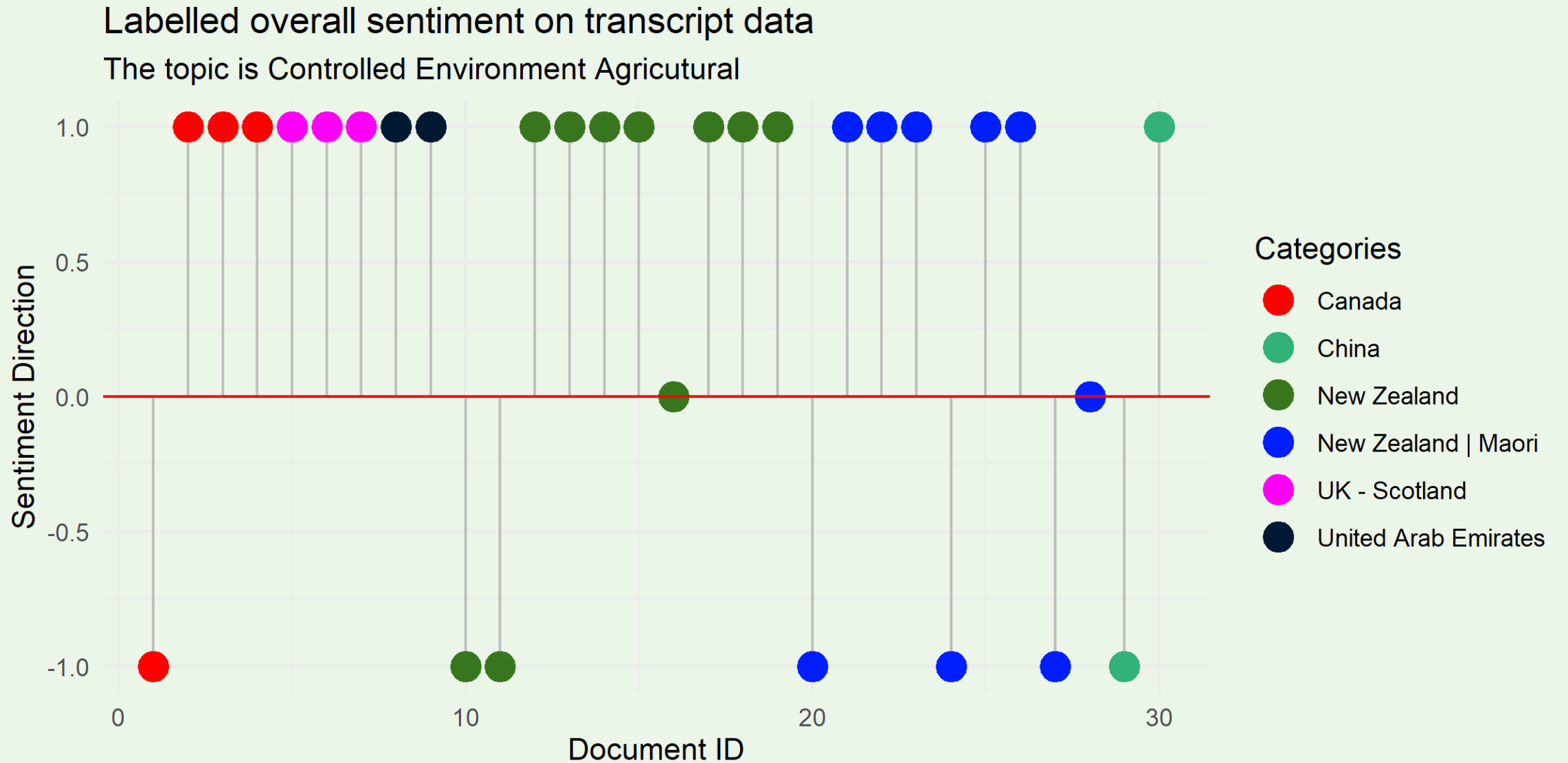
- » Sensitive information that can reveal personal identity
- » Time stamps
- » Abundance of stop words (words we use filling the gaps)
- » Mixture of the question and answers
  - » Not necessary to follow one question.
  - » Interviewer asks questions to clarify the question.

**Let's collaborate**

# Workflow



# Labelled Data





## Existing works – term frequency approach

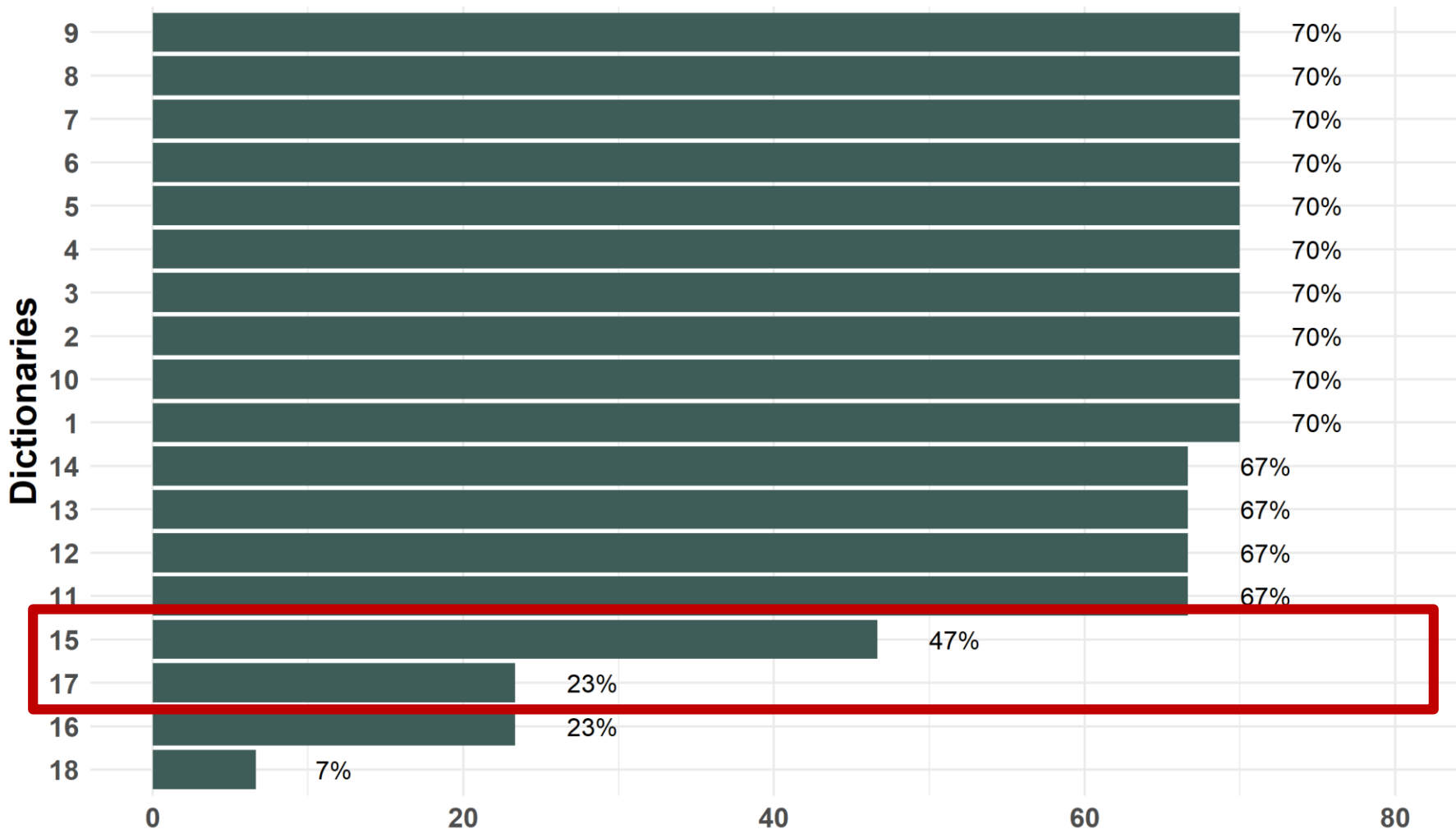
- » SentimentAnalysis – 4 built-in dictionaries
- » Syuzhet – 4 built-in dictionaries
- » Sentimentr – 10 built-in dictionaries
- » Sentence level scoring → aggregate to document level.



# Dictionary scoring

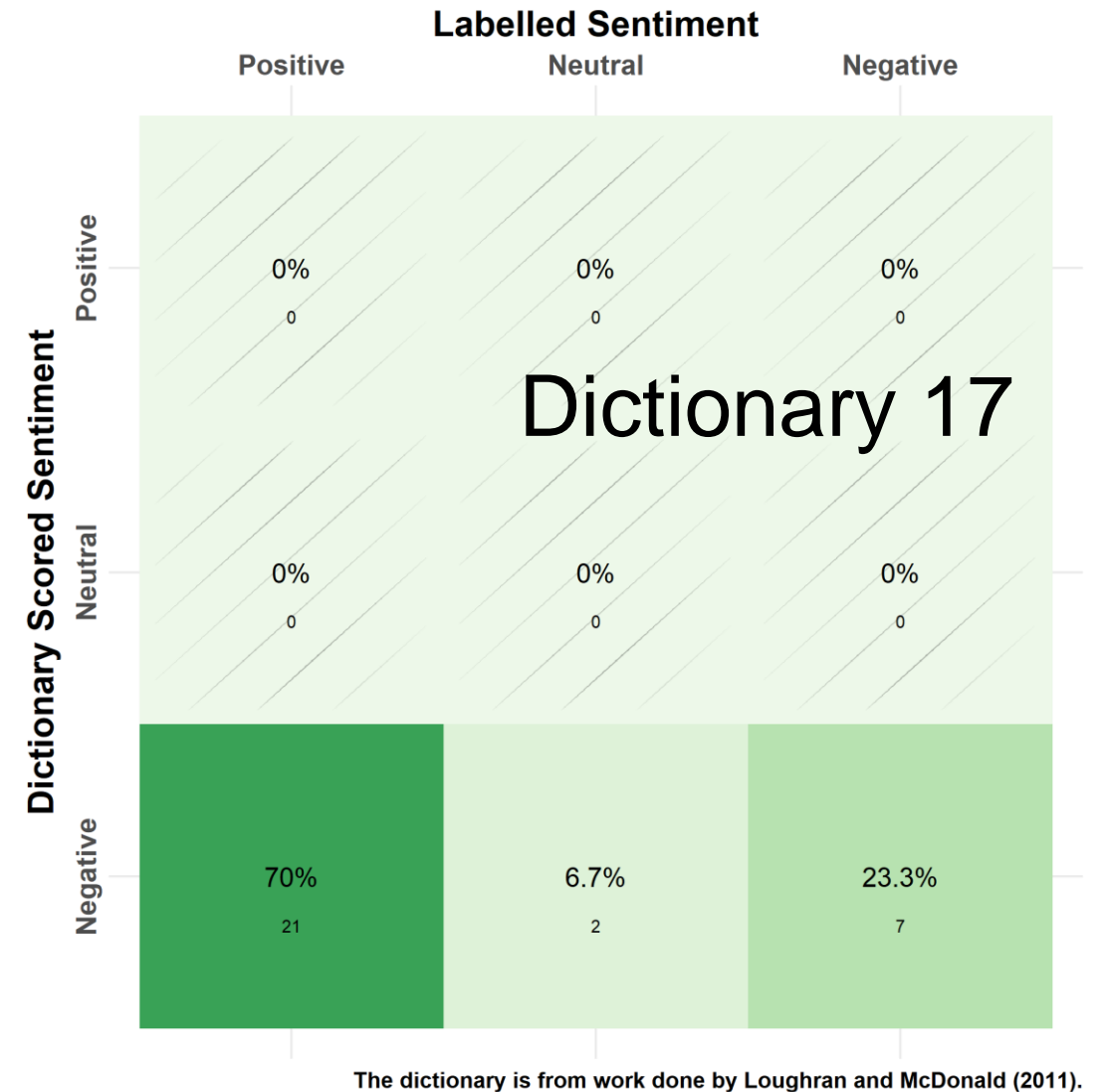
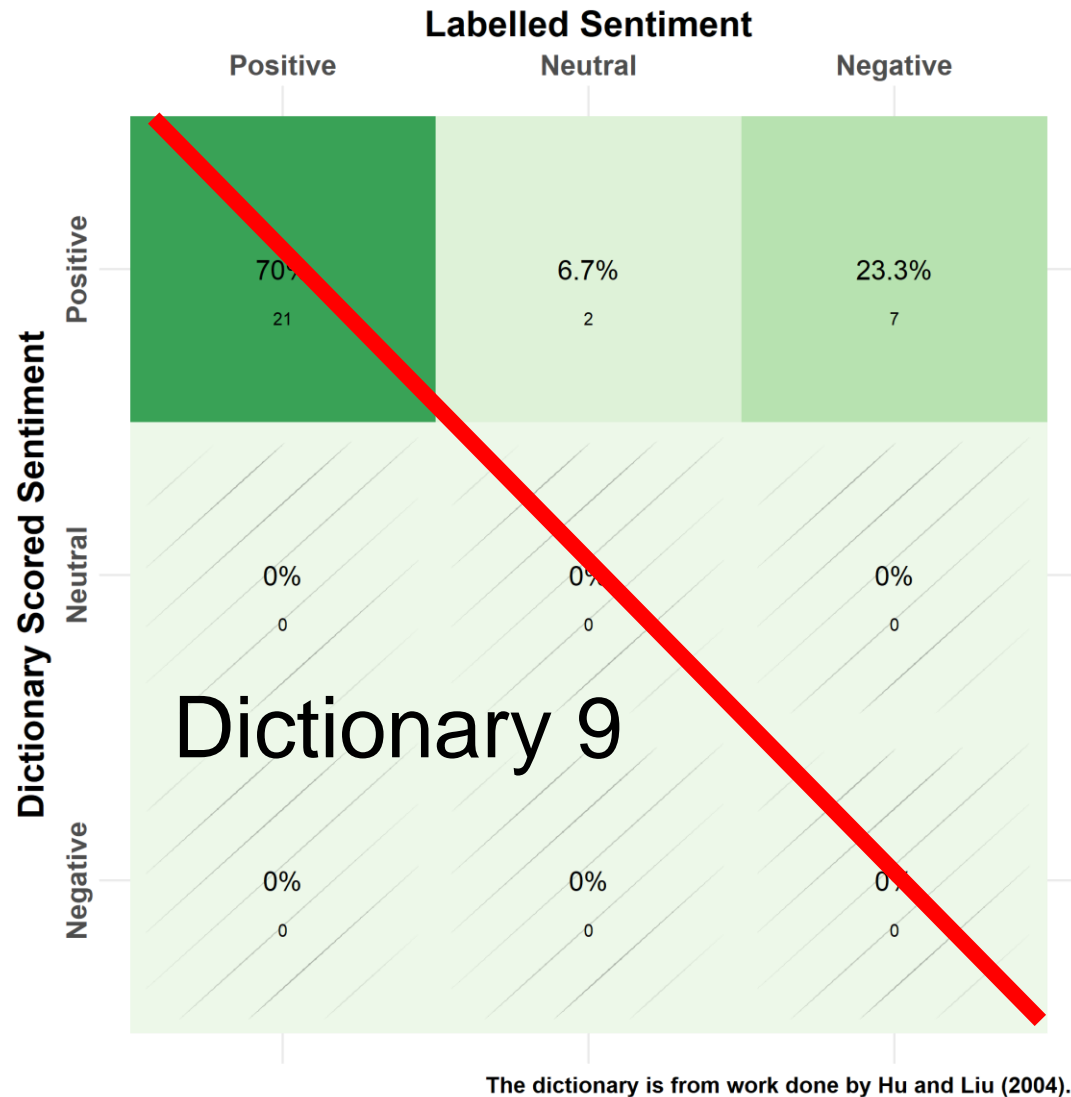


## Dictionary-based sentiment analysis on transcript data



**Scoring Accuracy (%)**  
Transcript data were preprocessed before the sentiment scoring.

# Acceptable vs Poor



## Recap dictionary based approach



- » 14 out of 18 built in dictionaries scored ~70% matched the labelled sentiment
- » Pre-built dictionaries are good at scoring positive sentiment
- » Same dictionary with different processing techniques yielded different scores



# Word representation – TF-IDF

## New Zealand industry

protect  
consent  
council  
customer  
gate  
unintended  
lend  
site  
cell  
fit

## New Zealand industry/scientist

overall  
versed  
domestic  
citrus  
welcome  
artificial  
commodity  
smart  
kiwifruit  
suitable

## New Zealand scientist

gene mutation  
hygiene bypass  
twig  
flower  
glasshouse  
sugar  
reliant  
biology

## Overseas industry

deliver  
consumption  
tray  
pollution  
equipment  
core  
vegetable  
basil  
client  
starter

## Overseas industry/scientist

advantage  
suitable  
strawberry  
quinoa  
potato  
glasshouse  
transportation  
affordable  
customer  
resistant

## Overseas scientist

greenhouse  
ordinary worker  
smart  
vegetable  
perception  
pollution  
facility  
controllable  
province

## New Zealand industry | Māori

methodology  
utilization  
teach  
macadamia  
hectare  
kumara  
hua  
kai  
manuka  
garden



## What's next

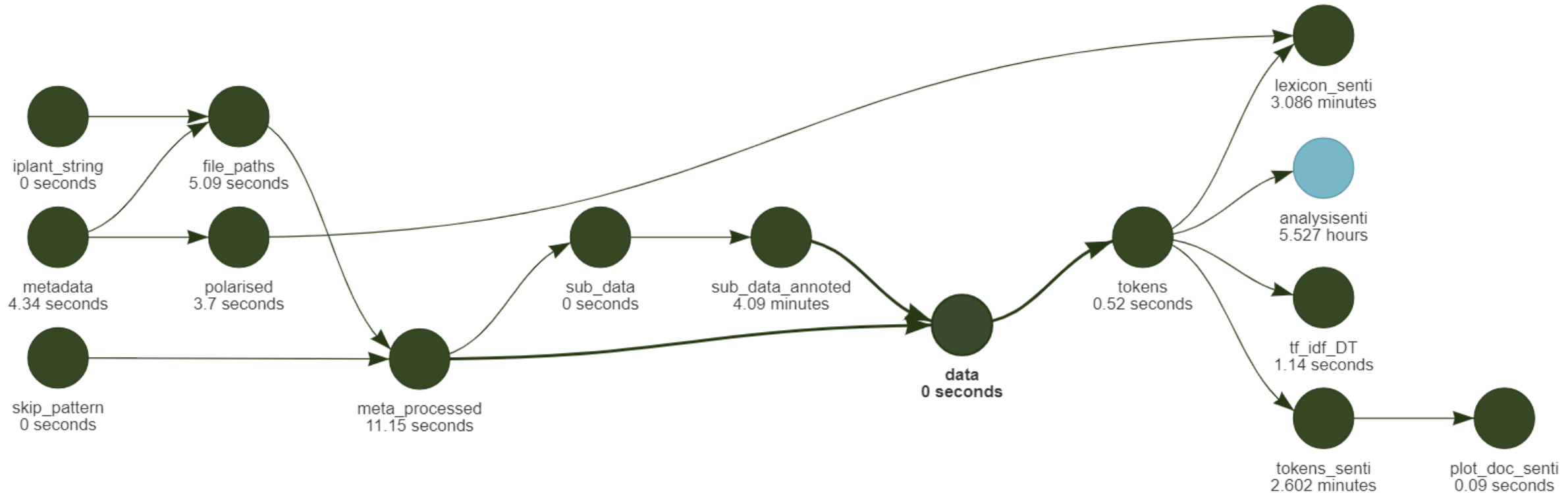


- » Dive into the sentence and word level analysis.
- » Develop context-specific sentiment dictionary.
- » Explore the application of word embedding in sentiment analysis.
- » Develop classification models.

# Tools for reproducibility



- renv
- usethis
- targets





**Jian.liu@plantandfood.co.nz**

plantandfood.co.nz

