



Beyond discrete genres: Mapping news items onto a multi-dimensional framework of genre cues

Zilin Lin¹, Kasper Welbers², Susan Vermeer¹, Damian Trilling¹

¹ University of Amsterdam

² Vrije Universiteit Amsterdam

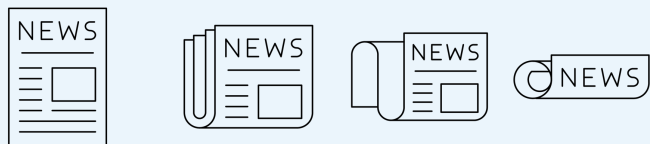
ICA 2022
30/05/2022

Email: z.lin@uva.nl
Twitter: [@linzilinn](https://twitter.com/linzilinn) | [@newsflows_erc](https://twitter.com/newsflows_erc)



How to study news?

Studying news before:



Studying news now:



Research question:

How can a vast and diverse amount of news items be mapped onto *one standardized framework*, representing the variety of what constitutes news?



Problems of the traditional genre approach



Organizing news items along abstract dimensions - Genres

Problem I: Discrete demarcations

- Top-down approach:
 - a pre-defined set of genre categories
- Bottom-up approach:
 - an open list of specific genres

**Solution I:
Moving beyond discrete genres**

Problem II: Newsroom-centered

- Over-reliance on the “supply-side” of journalistic practice
- Little evidence on the audience’s perception

**Solution II:
Studying genre cues**



Conceptualization & Operationalization

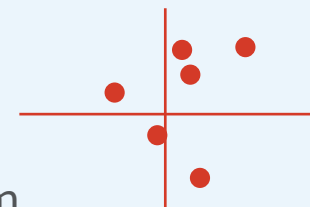
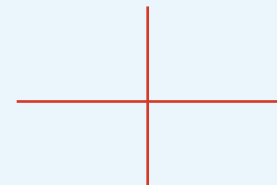


Conceptually, we propose:

- A **multi-dimensional framework** for studying **genre cues in a continuum**

We operationalize it by:

- Creating a **two-dimensional grid** as a starting point:
 - Linguistic features as the proxy of genre cues:
 - *Factuality & Formality*
- Automatically **positioning** individual news items onto this grid:
 - Training two sentence-level classifiers to predict news sentences
 - Calculating the proportion of factual and formal sentences for each item



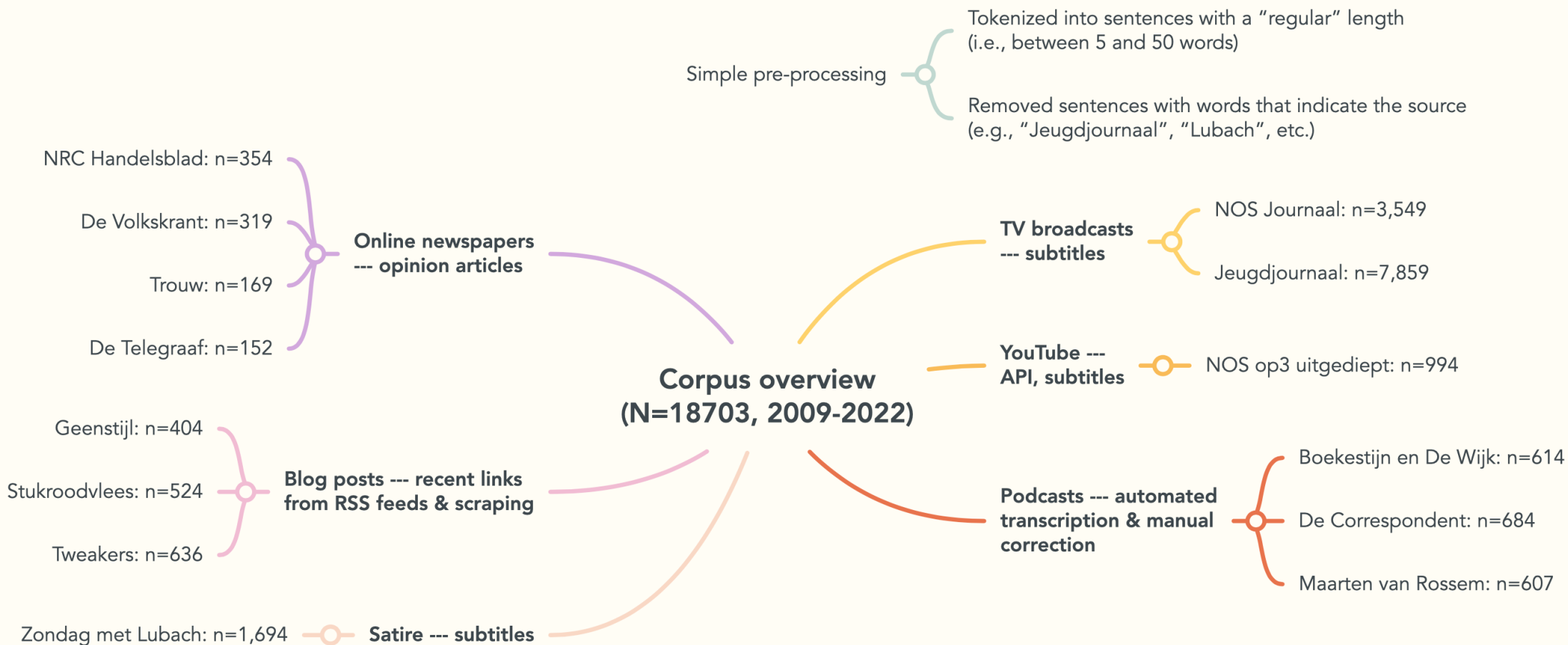


Method – Data acquisition and preparation



A/S
Go/R

NEWS
FLOW



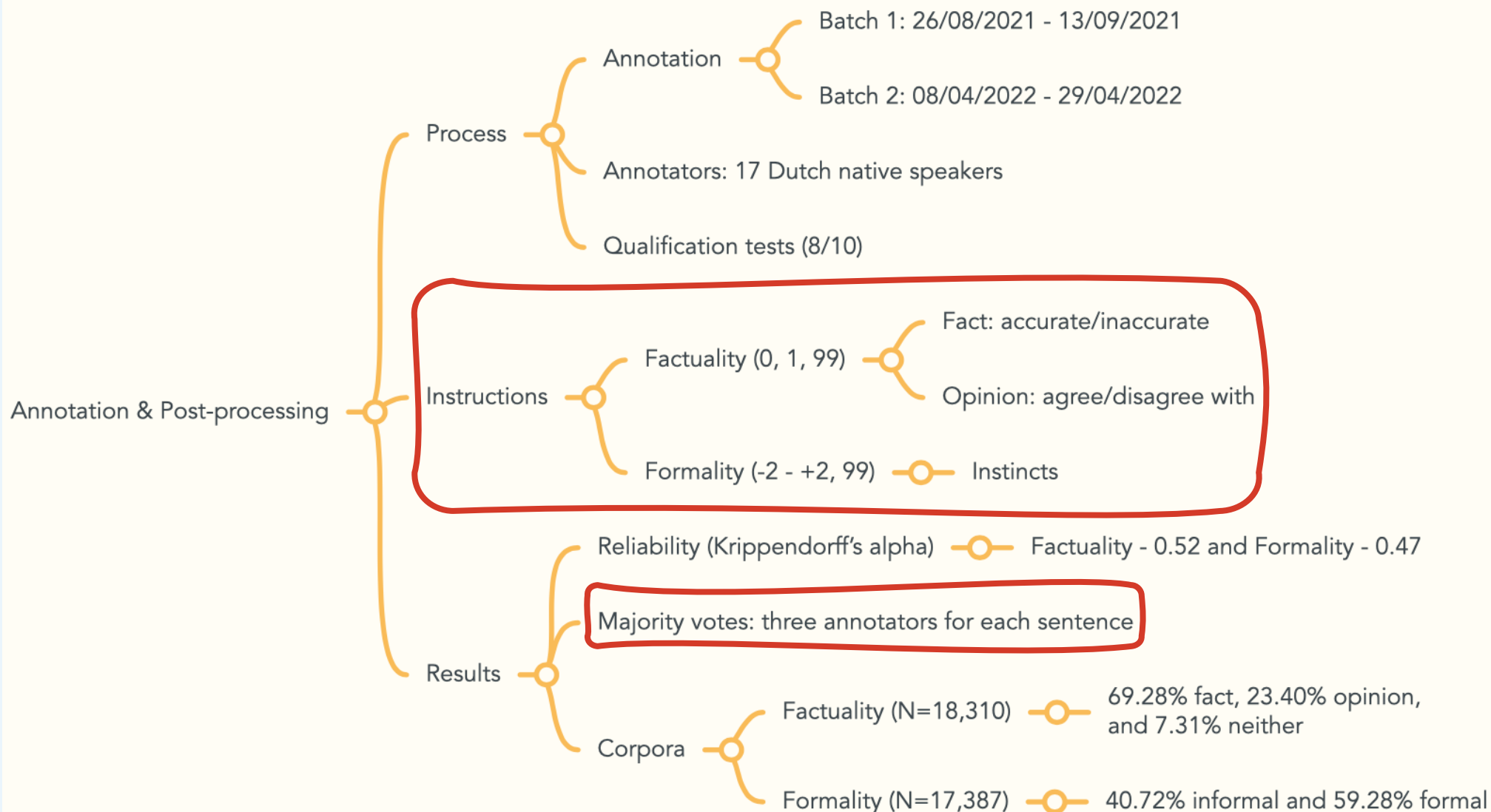


Method – Corpus construction



A/S
GoR

NEWS
FLOW



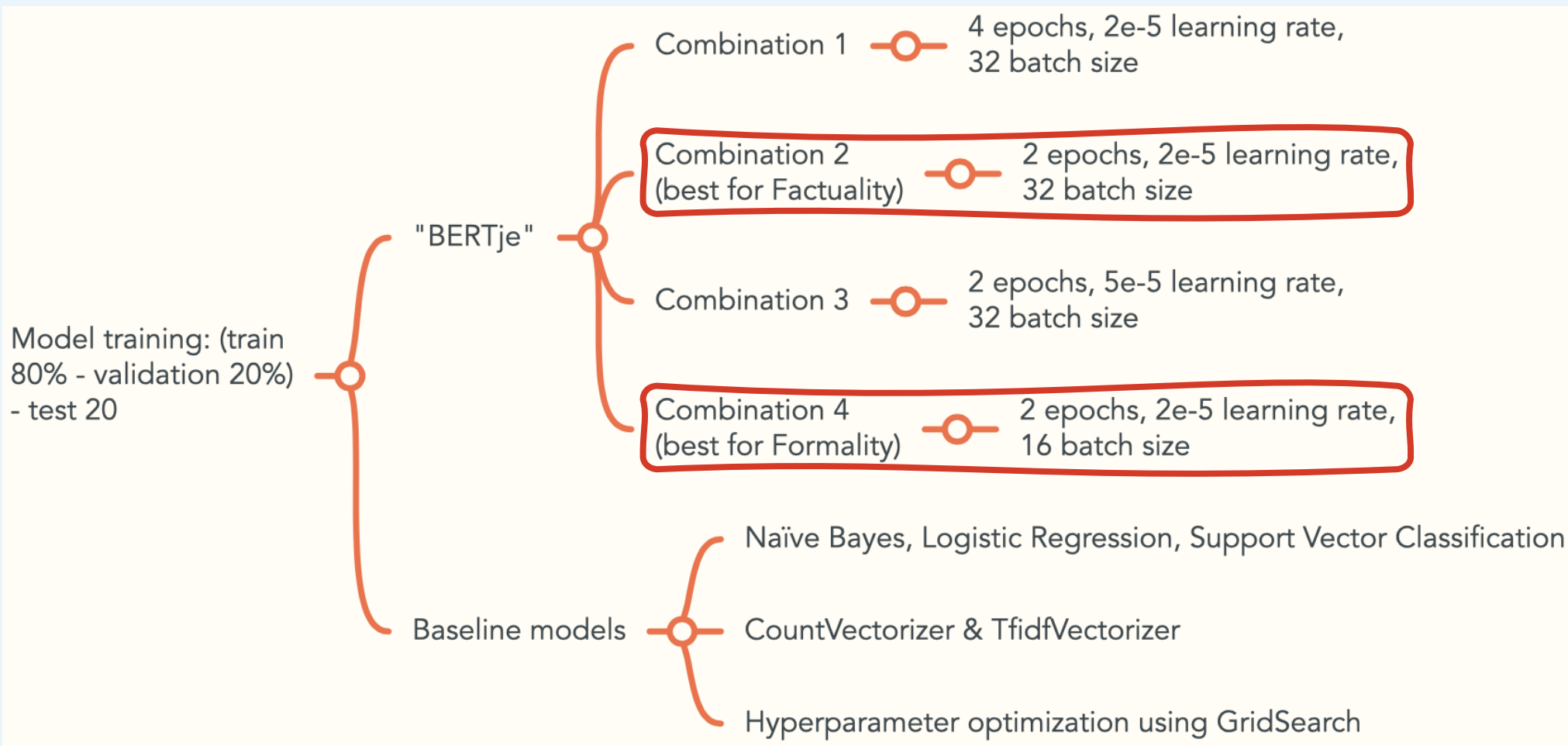


Method – Model training



A/S
GoR

NEWS
FLOW





Results – Model performance



A/S
GoR

NEWS
FLOW

Inormal - Formal

Model	F1-score (macro avg)
Naïve Bayes	0.78
Logistic Regression	0.78
Support Vector Classification	0.78
BERT	<u>0.86</u>

	precision	recall	f1-score	support
0	0.87	0.79	0.83	1393
1	0.87	0.92	0.89	2085
accuracy			0.87	3478
macro avg	0.87	0.85	0.86	3478
weighted avg	0.87	0.87	0.87	3478

Fact – Opinion – Neither of them

Model	F1-score (macro avg)
Naïve Bayes	0.50
Logistic Regression	0.51
Support Vector Classification	0.54
BERT	<u>0.79</u>

	precision	recall	f1-score	support
0	0.89	0.92	0.90	2524
1	0.76	0.70	0.73	856
2	0.78	0.72	0.75	282
accuracy			0.85	3662
macro avg	0.81	0.78	0.79	3662
weighted avg	0.85	0.85	0.85	3662



Conceptualization & Operationalization

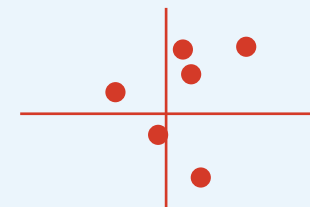
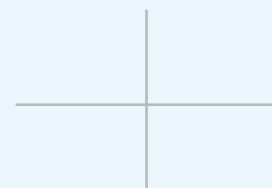


Conceptually, we propose:

- A multi-dimensional framework for studying genre cues in a continuum

We operationalize it by:

- Creating a two-dimensional grid as a starting point:
 - Factuality & Formality
- Automatically positioning individual news items onto this grid:
 - Training two sentence-level classifiers to predict news sentences
 - Calculating the proportion of factual and formal sentences for each item



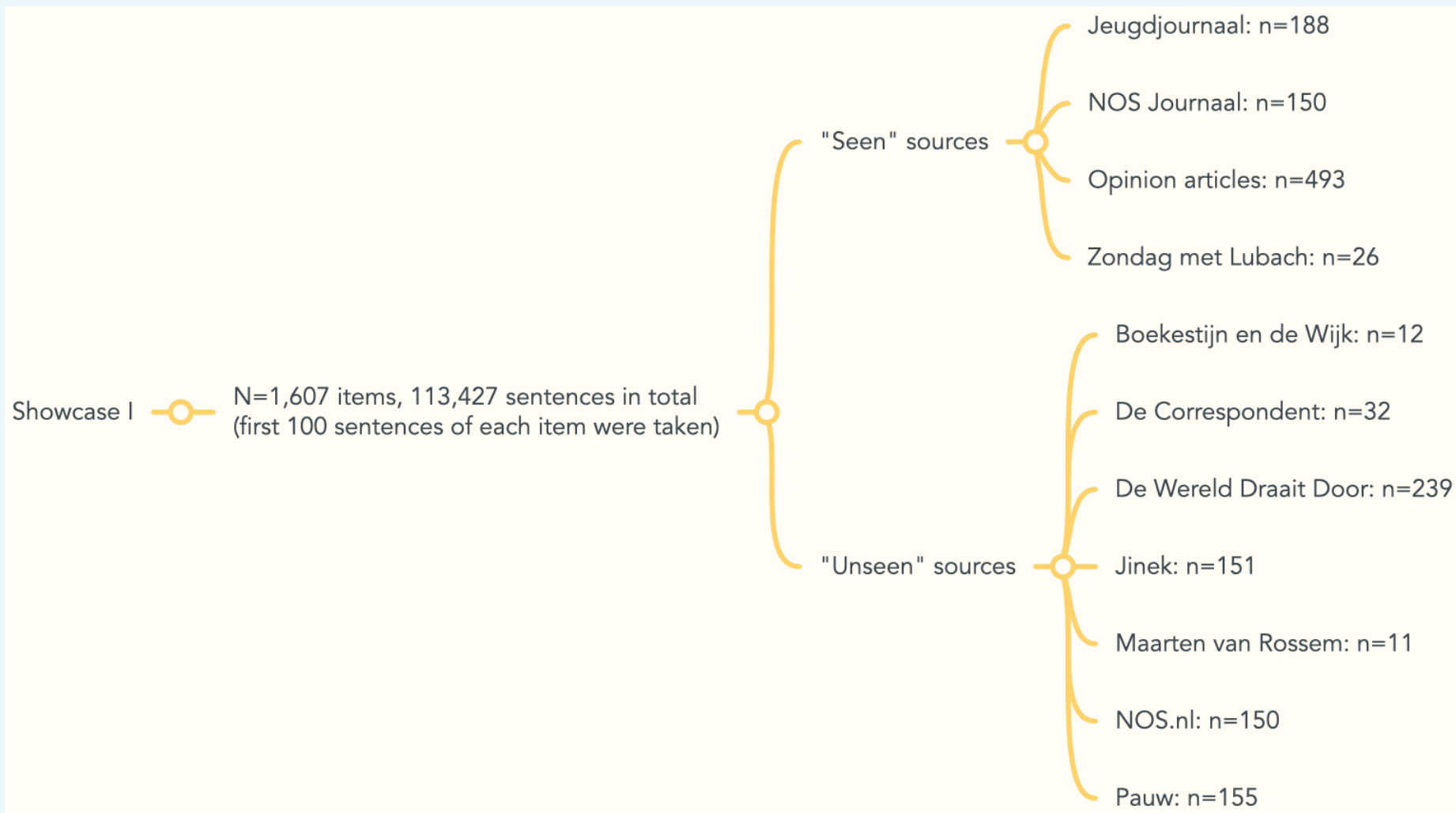


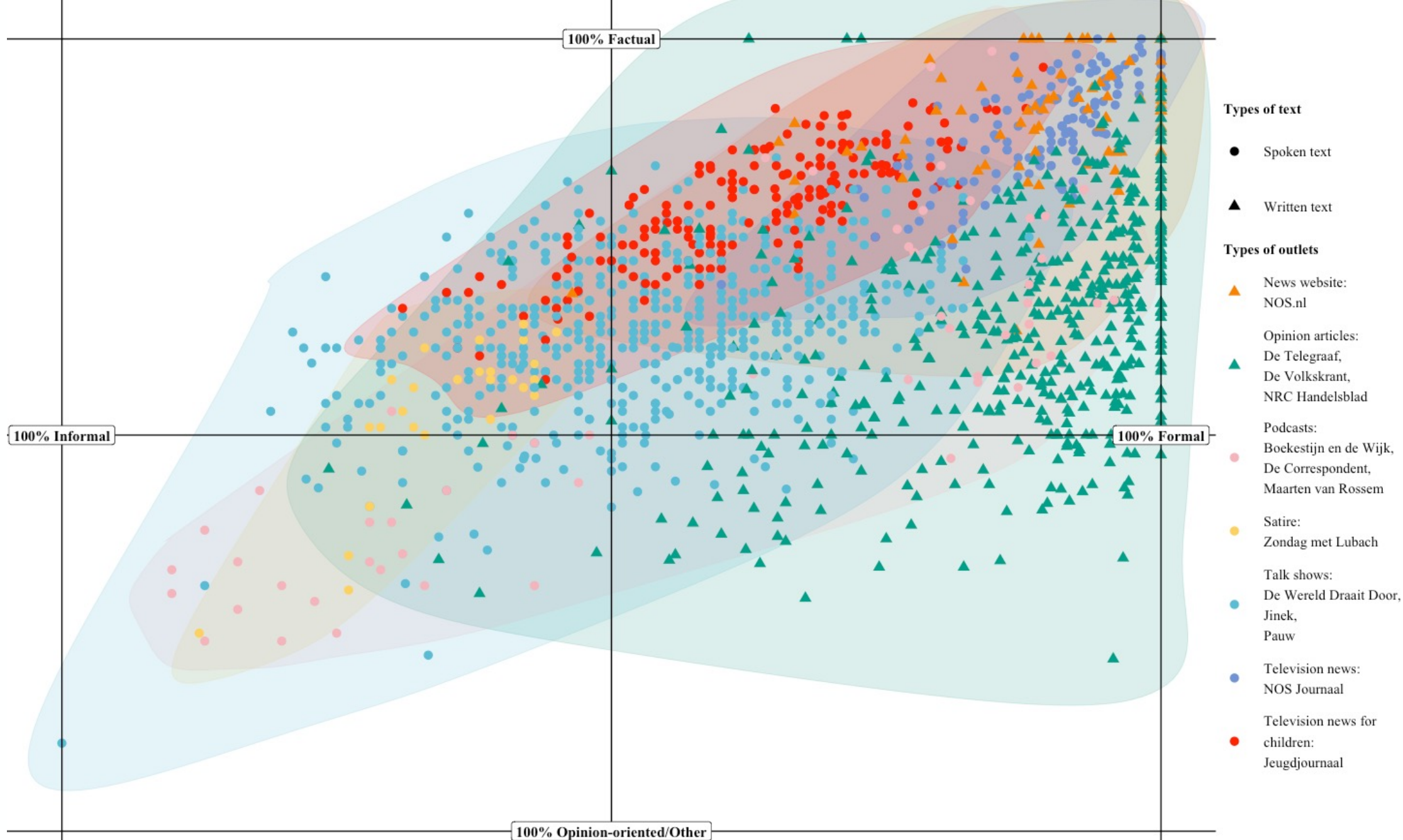
Results – Showcase I

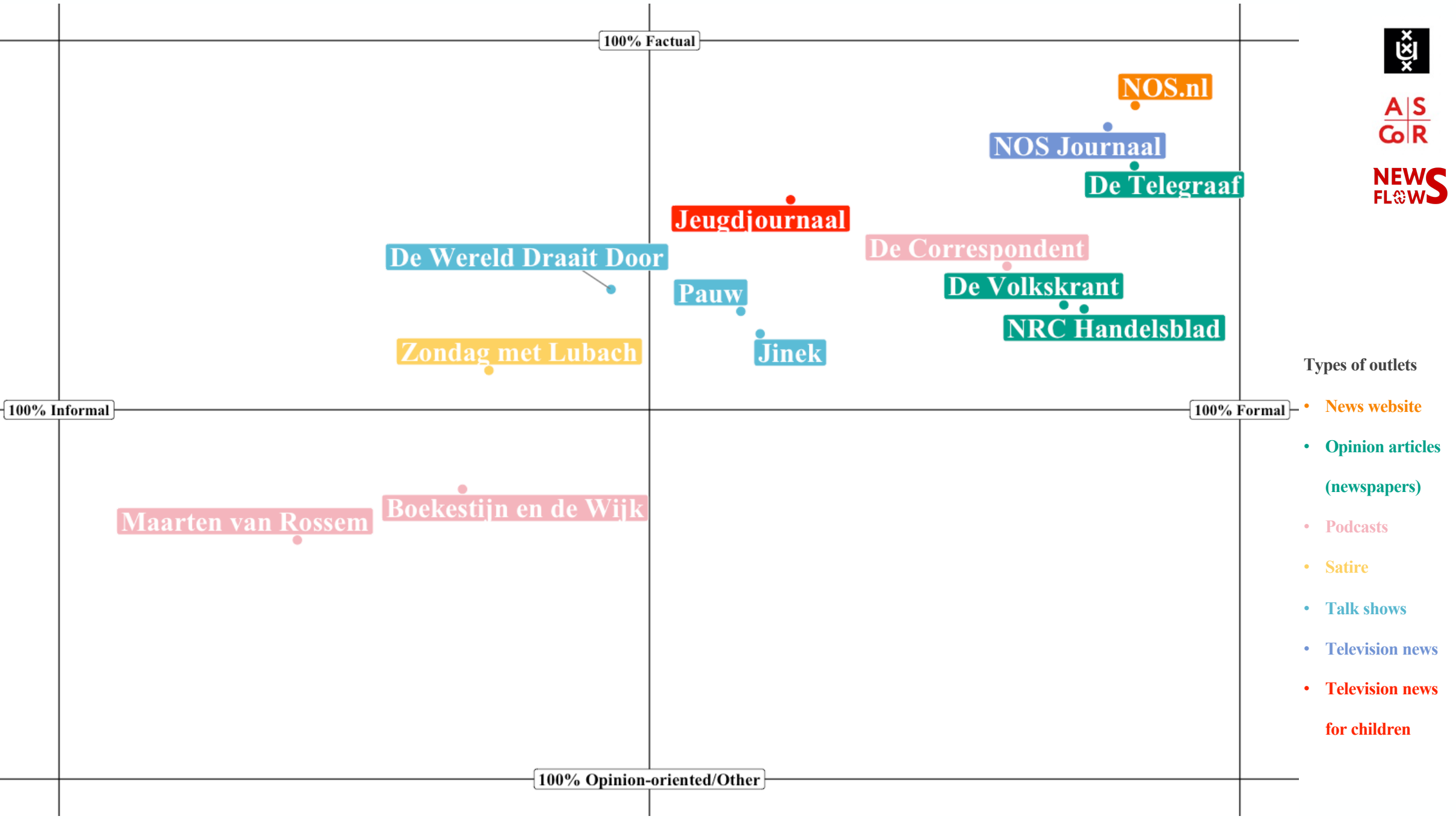


A/S
GoR

NEWS
FLOW

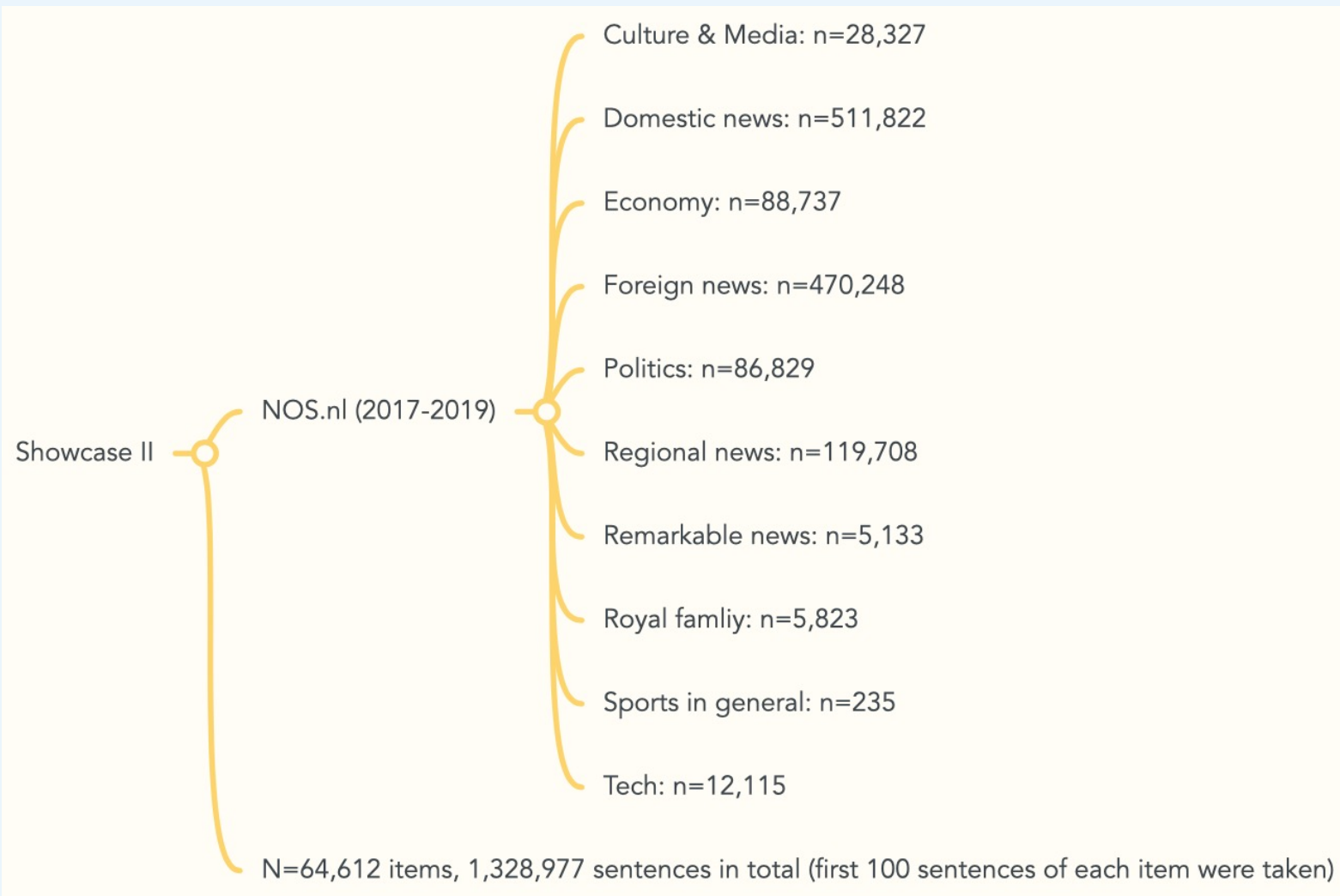


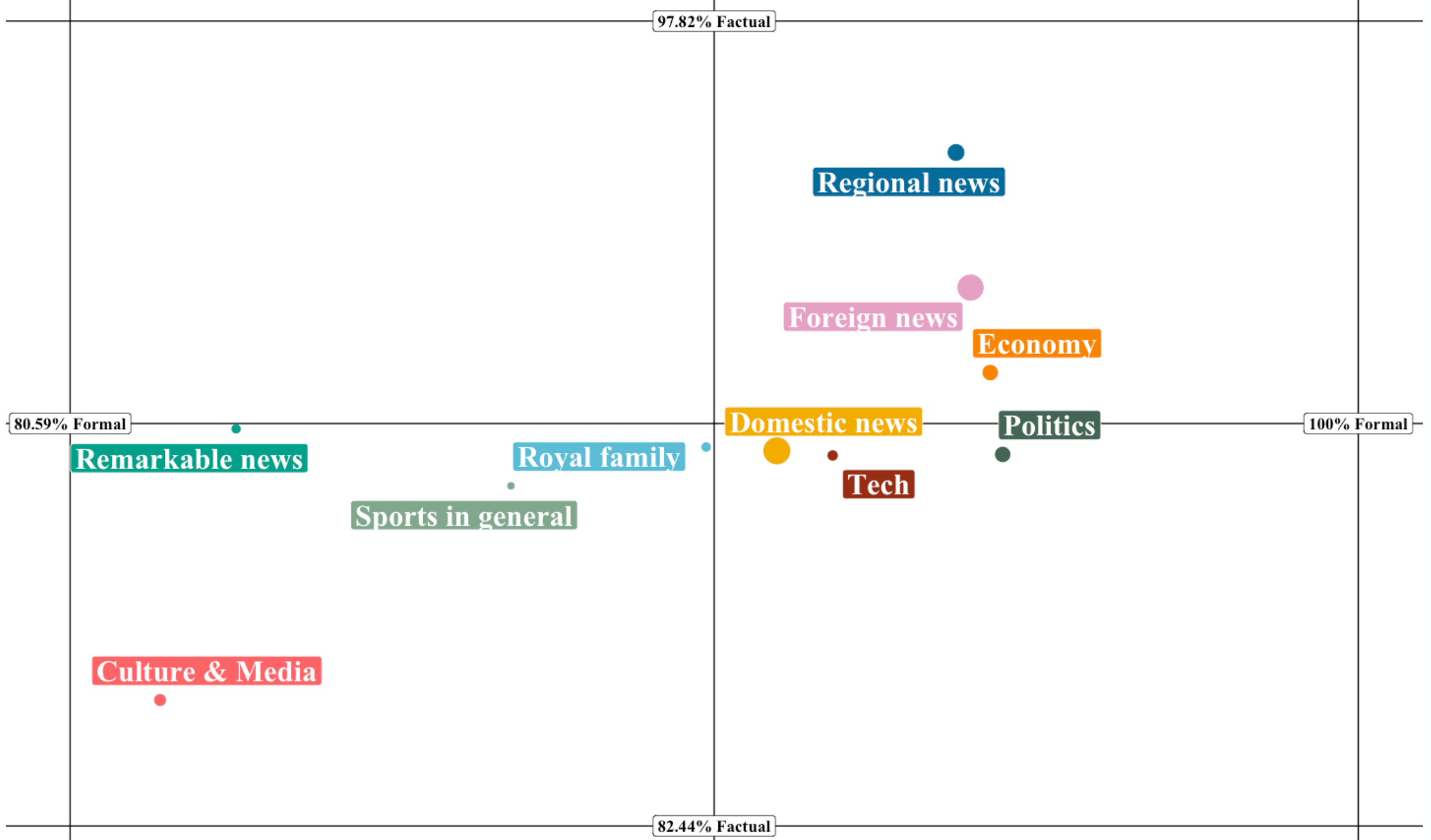






Results – Showcase II





A | S
Go | R

NEWS
FLOW

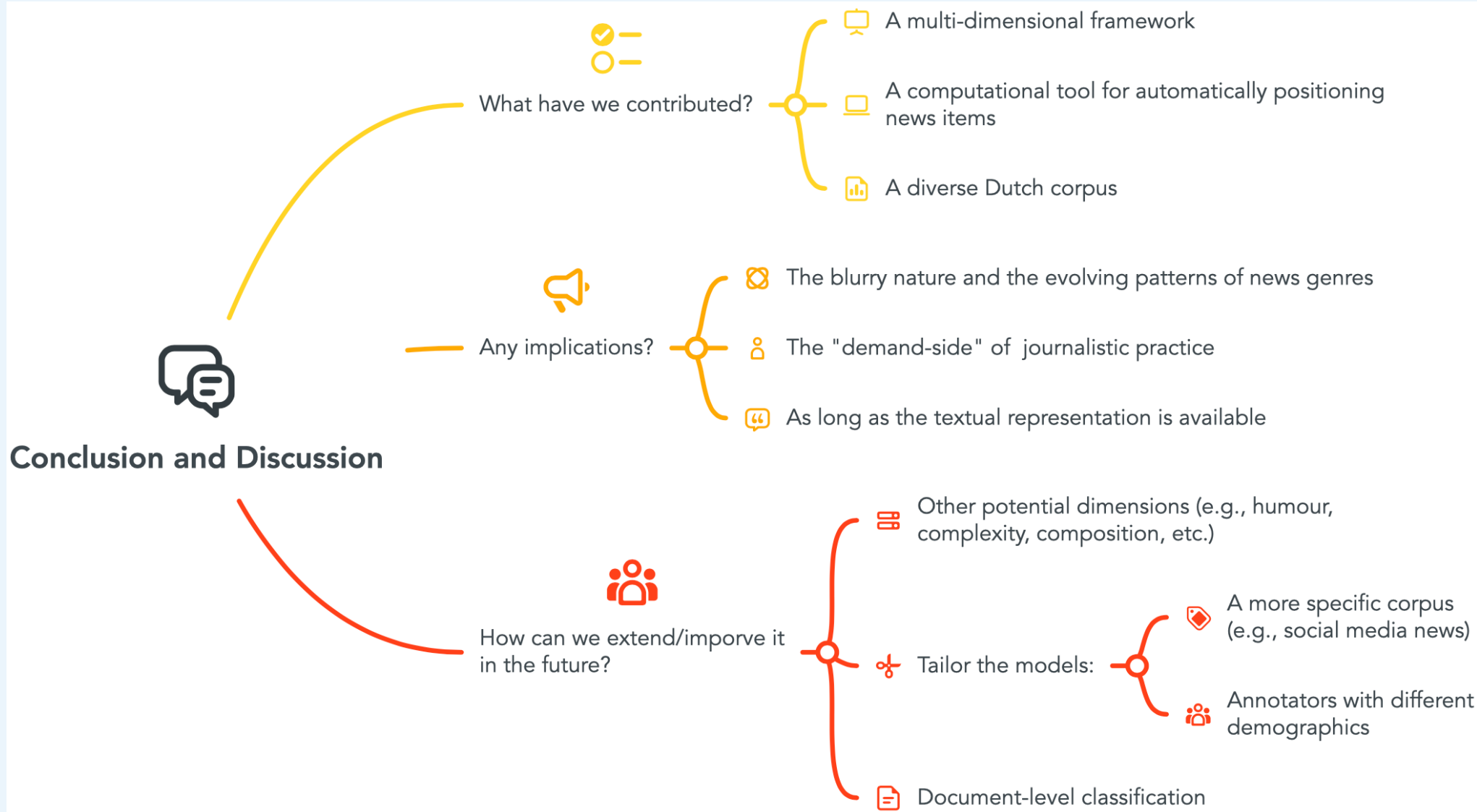


Conclusion and discussion



A | S
Go | R

NEWS
FLOW





Thank you!