Computational Literacy
Final Project
Nan Jiang

## Why Do We Still Believe in Milk?
## An Analysis of Propaganda Skills Adopted in Official News of Valio through Topic Modelling

1. **Introduction**

    The dairy industry is well-developed in Finland and it is a result of long-term propaganda (Maitomyytit-kampanjasta, 2022). Valio Group is a leader in the dairy industry as well as Finland's biggest exporter of food products. It not only produces milk-based and plant-based products but also engages in extensive outreach efforts for dairy food promotion (Reports and Financial Statements, 2022).

    From the standpoint of propaganda, the industrial promotion of dairy food is not only an enlightenment of public health but also leads to the increasing consumption of dairy food by informing and persuading the public sphere.

    In this research, I selected the famous milk brand in Finland-Valio as a breakthrough point to explore propaganda skills used in the dairy industrial campaign. The latest news content of Valio from 2017 to 2024 March will be investigated by computational analysis of Topic Modelling. The object of study is limited to the officially published news articles from Valio's website, which can avoid distractions from irrelevant factors as much as possible and make the outcome more representative within a specific boundary.

2. **Backgrounds and objectives**
    a)  Milk Myths and Milk War

    The propaganda of milk is vital in the advertising of dairy product brands and communications of the government. "The Finnish Milk Propaganda Bureau" was founded in 1929 by milk producers as well as the government. A variety of enlightenment on the advantages of milk was conducted for the public, especially at schools targeting children. Campaigns promoting the positive impact of milk-based products on health, such as cream and butter often extoll the virtues of milk over coffee. After World War II, Milk was strongly subsidized (in Nordics and internationally) as an effective means to fight

undernourishment (Jallinoja, 2020).

However, a counter-propaganda of milk happened in 2020, named Oatly's "Milk Myths" campaign, in which a Swedish oat product brand "Oatly" sent a "Milk Myths" booklet to the homes of Finnish schoolchildren. The booklet presented 20 statements called milk myths related to the production, consumption and nutritional value of milk, and Oatly refuted them in the booklet. What's more, Oatly's representatives stressed that the campaign is also concerned about issues of climate change (Jallinoja, 2020).

In summary, the nutritional value of dairy products is the core question in this "milk war". Animal fat is suspected to be harmful to people's health, which has led to more people choosing vegetable fat instead. Besides, the negative impact on the global climate of cattle farming is also a point that the dairy industry is being accused of.

b)  Propaganda skills

Propaganda is a designed activity, through which the controlling people or organization reorganize specific information and sends it to targeted receivers in a systematic method to change their opinions and actions in the end. It is quite common in the marketing field with an original intention to persuade customers to buy specific products through promoting the goodness of it. (Snow, 2019, p.1).

There are two typical techniques which are supposed to be used in the Finland milk industry. The first is Glittering Generalities, which connects the product with shining ideas to appeal to our emotions of love(Clyde R, 1937). Previous research found that health, environment, economy, and Finnishness are the most common topics in the comments about milk-based food from related experts and news (Maitomyytit-kampanjasta, 2022).

Another is Testimonial, which means statements from prominent people are promoted with the expectation that the crowd will follow the leader(Clyde R, 1937). In most of the reports of Valio, there are lots of quotations from experts or government departments, which back up their advertising of the goodness of dairy products.

However, in the context of the Milk War, whether the dairy industry adopts different techniques of propaganda to fight against the opposite opinions is the jumping-off point of this study.

c)  Computational methods

I plan to use Topic modelling (TM) to capture the main topics in the news articles of Valio. TM is a family of computational algorithms that aim at "discovering the main themes that pervade a large and otherwise unstructured collection of documents". Once I have an understanding of the content of the news, I can conduct a propaganda analysis about how Valio promotes milk-based food consumption in changing times through public communication.

d)  Research questions

Q1: What are the main themes in Valio published news from 2017 to 2024?

Q2: What kinds of propaganda skills are buried in the public news of Valio?

## 3.  Data retrieval and preprocessing

My dataset is a news archive of Valio from 2017 to 2024 March. https://www.valio.com/news-archive/. Public credibility is crucial when communicating serious food-related knowledge. Besides, all kinds of news will use the news published in official sources as a basis. Hence, I chose the news collection on the official websites of Valio to be data material for content analysis.

I used Python to retrieve raw data and preprocess it. I first fetched the text content of each link on the webpage by web crawling and only kept the title and main body of each article and finally converted it into a JSON file. This process is stored in *webcrawling2.0.py* and *valio21.json.* Then I processed the outcome of the last step into wordlists, removed the stopwords, punctuations and numbers, and fed it into the LDA model. The outcome of topic modelling was visualized later and exhibited in an Intertopic Distance Map. This process is stored in *LDAmodel_Visualization.py* and *lda_visualization21.html.*

## 4.  Methods and implementation

Topic modelling (TM) was conducted to figure out the main themes in the information Valio conveyed to the consumers. I created a LDA Model with Gensim, and transformed the model's data for visualization by pyLDAvis. I set the number of topics to be 10 and 5 successively. When I observed the Intertopic Distance Map

of 10 topics, I found that 5 of them have a small marginal topic distribution. So I supposed that some less prevalent topics could be combined. When I changed the topic number to 5, the distribution of topics was more readable without any overlap though topic 3 still has a small marginal topic distribution. The overall situation was much better than before. Therefore, I decided to use the outcome of 5 topics to conduct the final analysis:

Topic 1: 0.034*"valio" + 0.010*"milk" + 0.009*"new" + 0.009*"dairy" + 0.008*"energy" + 0.007*"cooperative" + 0.007*"osuuskunta" + 0.007*"ice" + 0.007*"children" + 0.006*"finland"

Topic 2: 0.025*"valio" + 0.015*"biogas" + 0.012*"plant" + 0.010*"milk" + 0.010*"packaging" + 0.009*"plastic" + 0.008*"new" + 0.007*"production" + 0.007*"finland" + 0.007*"manure"

Topic 3: 0.007*"plantbased" + 0.006*"finland" + 0.006*"berry" + 0.006*"gable" + 0.006*"industry" + 0.006*"valio" + 0.005*"project" + 0.005*"fairtrade" + 0.005*"pilot" + 0.005*"criteria"

Topic 4: 0.041*"valio" + 0.017*"milk" + 0.016*"products" + 0.015*"food" + 0.009*"dairy" + 0.007*"new" + 0.006*"sales" + 0.006*"market" + 0.005*"product" + 0.005*"business"

Topic 5: 0.022*"valio" + 0.017*"milk" + 0.016*"carbon" + 0.012*"dairy" + 0.012*"farms" + 0.011*"production" + 0.009*"methane" + 0.008*"farmer" + 0.008*"emissions" + 0.008*"footprint"

5. **Results analysis**

**Q1: What are the main themes in Valio published news from 2017 to 2024 March?**

To get more information from the word list, I expanded my observation to the top 30 words shown in the Intertopic Distance Map to do the extraction of topics.

The topic 4 is the most significant one. "Product" and "products" both appear in the top 10 words. When I expanded the word list to the top-30, I found more terms regarding the features of products, like "ingredients", "lactose", "protein", "sugar",

"oddlygood" and "taste". Lactose-free is a milestone in the technology development of Valio, and these words all show that Research&Production remains a core content in Valio's promotion.

The second significant topic is topic 5. There are notably interesting nouns- "carbon", "footprint", "climate", "sustainability" and "welfare". This group of words could refer to Carbon footprint calculation, which is an essential concept related to responsible and sustainable operations. Besides, another important group of words are "farms", "methane", "cows", and "animal". "Methane" refers to a main coproduct in cattle farms and Valio used to be accused of the process of farming which produced lots of methane. Based on these two groups of words, Valio may tend to communicate its actions about the carbon footprint like solving the problems of methane and other coproducts of dairy farming.

The third significant topic is topic 2. There are two groups of keywords. One includes "biogas" with "plant", "manure" and "renewable", and the other is "packaging" with "plastic", "materials", and "recycled". "Biogas" is a type of renewable energy source produced by agricultural by-products, while "plant" and "manure" implicate the recycling of cattle's manure for planting. Another key group of words is about "packaging", which can refer to that Valio is recycling the product packaging and reducing the use of plastic material. In summary, this topic is related to environmental protection as well in terms of the by-products and packaging.

In topic 1, "cooperative" and its Finnish word "osuuskunta" stand out, along with "children", "sports", and "factory". These words could indicate a new production mode, which includes more cooperation with local people.

In the least significant topic 3, there are more surprising but scattered words, like "plantbased", "berry", "fairtrade", "pilot", "criteria", "world" and "international" along with "project", "new" and "way". It could be speculated that this topic is about international trade with high-level business criteria, which shows that Valio is working on expanding the market worldwide. Moreover, it can also be a new project about fair trade regarding berry or other plant production.

Regarding topics 1 and 3, the meaning of words is blurred and unfocused, which calls for additional close reading of the news. After checking several articles in the news archive, I found there exists some relevant content which can explain the two

topics to a certain extent. For topic 3, Valio joins an innovative project about dealing with the human rights challenges in the berry-picking industry, which is initiated by Fairtrade. For topic 1, the most relevant news is about Valio Oy's General Meeting, in which the Supervisory Board positions were divided among the owner-cooperatives. This action shows Valio attaches great importance to cooperation with local farmers.

Besides, two points need to be paid attention to. First, "finnish" and "finland" appear in every topic-30 word list even though they are not listed in the top 10, which shows the concept of the nation plays a vital role in the news. Second, "farmer" or "farm" appears in topics 1, 2 and 5. It can indicate Valio attaches great importance to the demonstration of cooperation with farmers.

**Q2: What kinds of propaganda skills are buried in the public news of Valio?**

To check the reliability of the interpretation of some themes generated from TM, I looked through the titles of these news articles. And there are related articles, which provide logical evidence for the interpretation of the topics. Then I conducted an analysis of possible propaganda skills of Valio based on the themes found :

a) Glittering Generalities & Predispositions of the Audience: Creating Resonance

Glittering Generalities is a device by which the propagandist identifies his program with virtue by use of "virtue words". And if the message is produced from the existing shared interests and beliefs, it will be more likely to win people's agreement, which is called creating resonance (Jowett & O'Donnell, 2015).

The topics 2 and 5 are both about environment-friendly issues, covering packaging, methane processing and manure reuse. The environmental protection theme has earned public attention worldwide. It is a critical concern in building and exhibiting a good brand image and a company's social responsibility. Hence, propagandizing the operations in the area of environmental protection is beneficial for Valio to earn a good reputation and win public advocacy.

What's more, the promotion of innovation in R&P shown in topic 4 is a

shining idea as well which is attractive to the public and creates a more direct connection with people's daily consumption. Besides, participating in the Fairtrade project about human rights is also an action which is supposed to win public favour.

b) Card stacking

However, the primary focus of milk promotion before was the nutritive value. Currently, environmental protection and other social issues have become the most important themes discussed in the Valio official news.

Given the background of "Milk War", it is known that questioning the nutrition of dairy food is the main content in the statements of "Oatly", and there is more and more non-supportive evidence for dairy food from experts. Even Valio itself creates a new branch of plant-based products—"Oddlygood". Therefore, Valio initiatively changes the attention of the public to their endeavour in the field of environment protection rather than gets stuck in the debate of dairy food nutrition. Recently, its enrolment in the Fairtrade project supporting human rights in the berry-picking industry can also benefit its public reputation even though it is not directly related to its own production.

This action is a skill of Card stacking in propaganda, which is employed by propagandists to under-emphasize and over-emphasize to dodge issues and evade facts(Clyde R, 1937).

c) The Transfer Device:

The transfer is a device by which the propagandist carries over the authority, sanction, and prestige of something people respect and revere to something he would have us accept(Clyde R, 1937). The Finnishness is important in the news content as "finnish" and "finland" have a high weight in the article. Thereby Valio transfers people's trust and support for the nation to their company's campaign or program.

d) The Plain Folks device

The Plain Folks device is used by politicians, labour leaders, businessmen, and even ministers and educators to win public confidence by appearing to be common people like people themselves (Clyde R, 1937). Valio shows local people have great engagement in its business by having representatives from

cooperatives on its Supervisory Board. Additionally, "Farmers" and "farms" are frequently mentioned in the news, which shows Valio cares about local interests and has a good relationship with local people. Staying close to the primary supporters embodies the Plain Folks device.

6. **Biases and limitations**

The news archive only shows limited content in the communication of Valio dairy products. For example, a large amount of social media advertising is lost.

Also, as an official source of brand information, it acts as first-hand information to generate content in other communication activities, but it may lack the narrative form and content which are designed specifically for consumers for better effectiveness of persuasion. So the propaganda skills should be explained within the context of official sources.

As for the topic modelling, its limitations are obvious along with its strengths in the procedure of research. First of all, it can extract the most common groups of words from a sea of documents, which acts as an important clue for figuring out the core themes. However, close reading cannot be overlooked in this distant reading process. I indeed checked the content of some news articles to figure out the blurred meaning of the word list in topics 1 and 3, and to evaluate the interpretation of the topic based on the word-list outcome.

On the other hand, according to the research before (Brookes & McEnery, 2019), topic modelling has limited use for digging out the topics and is prone to generating mixed results with non-existing themes. Given my previous trial with problematic coding, it is likely to make false inferences if researchers manually assign meanings to those topics by combining the separated words into a whole story without a sound theoretical framework. A severe problem is if there is no such "story", researchers just waste their time making up a topic from nothing.

Similarly, another subjective factor existing in topic modelling is the number of topics, which is chosen by humans, though the whole process is done by algorithm objectively (Brookes & McEnery, 2019). It leads to the cherry-picking problem from the "try and see" approach. Besides, Brookes & McEnery (2019) illustrated another inescapable drawback of Topic modelling is that it is a non-deterministic method. It is not a reproducible process, which makes the human-supervised

interpretation more flexible and less reliable as others cannot cross-check the outcome. In my case, it is possible to get different themes in different trials.

Moreover, I did not conduct lexicon normalization like stemming and lemmatizing in the preprocessing of the words. On one hand, the dataset is not that big and I want to keep the content as detailed as possible. On the other hand, I would like to see "farm" and "farmers" as two kinds of information even though they share one lemma. However, I am not sure if it will impact the outcome of the LDA model. Furthermore, I removed numbers in the text, but the utilization of numbers can be a way to make the content look more professional and authentic. The possible skills related to numbers may be hard to verify through the methods designed for analyzing text data. These two operations reflect one of the vital concerns about topic modelling—the lack of linguistic sensitivity. The removal of stop-words, the stemming and the ignoration of the order of words all contribute to disregarding syntactic and grammatical contexts of word use (Brookes & McEnery, 2019). These factors contain lots of clues about the construction of the themes based on various meanings of words within a document.

## 7. Conclusion and further thoughts

a) In summary, the main themes and features of news content on the official websites of Valio from 2017 to 2024 March are discovered through Topic Modelling. The core communication content of Valio is converted to product innovation and social issues like environmental protection and human rights from the nutritive value of dairy products under the circumstances of Milk War. Besides, Finnishness and highlighting its cooperation with farmers are two prominent features. In terms of propaganda skills, Glittering generalities, Creating resonance, Card stacking, Transfer device, and Plain folks device are detected through the analysis of the news content.

b) In future studies, the dataset can be expanded to the content about dairy food from different types of media and subjects, such as social media and government reports. Moreover, the effect of propaganda can be measured by analysing the reflection of the targeted audience, for which the dataset can be the collection of consumers' comments from media or e-commerce platforms.

In terms of the method, a possible improvement is to add the function word into the analysis. According to Brookes & McEnery( 2019), even punctuation

can have a role in the domain of linguistic discourse analysis. For example, the propaganda skill of the Testimonial device often comes about by the company of quotations from experts. This skill is observed through the initial close reading but was not shown in the topic modelling outcome since the punctuations were removed. Hence, involving more linguistic theory can help to give rise to a more reliable and comprehensive process for discovering topics. Moreover, to provide a more scientific method in the digital humanity domain, how computational methods like topic modelling could be blended with human analysis more systematically is worthy of further investigation.

## 8. Bibliography

Brookes, G., & McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies, 21*(1), 3–21. https://doi.org/10.1177/1461445618814032

Clyde R, M. (1937, November). How to detect propaganda. *Propaganda Analysis,* 1.

Jallinoja, P. (2020). Maitosotien jatkosota. *Politiikasta. fi*.

Jowett, G. S., & O'Donnell, V. (2015). How to Analyze Propaganda. In Propaganda & Persuasion (6th ed). Sage.

Lavikka, T. (2008). Evaluating Differences in Desired Brand Image in Two Markets– Case: Valio Lactose Free Products in Finland and Sweden.

Lukkarinen, V. (2022). Kasvijuomat horjuttamassa suomalaisten maitouskoa: Diskurssianalyysi Maitomyytit-kampanjasta.

Phan, T. (2020). A Market Expansion Research for F&B Distribution Agents in Ho Chi Minh city, Vietnam: Importing Valio Oddly Good® through Brand Image.

Reports and Financial Statements. (n.d.). valio.com. Retrieved December 20, 2023, from https://www.valio.com/sustainability/reports-and-financial-statements/

Sattarova, A. (2016). Marketing the Eco-Efficiency of a Finnish Product in Russia.