AUTOMATED INTERPRETATION
THE OTHER, BETTER, WISER AI

- Machine Learning models often claim to predict rules by analysing data and answers.
- This is empirical analysis to an extreme, but risks omitting extraneous data created from unpredictable factors like human behaviour.

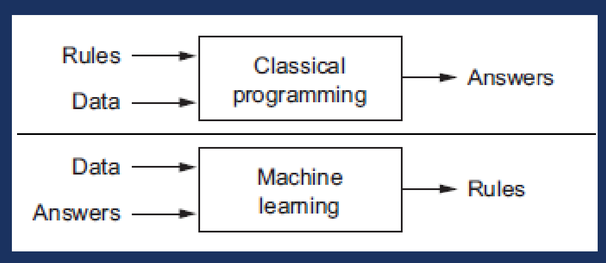

Figure 1 – A comparative model of procedural programming and machine learning.

- By observing seminars as case studies, and evaluating them as qualitative data, it was possible to gain insight into socio-industrial trends and methods through the perspective of workers within that context.
- These seminars can offer a large set of rich qualitative data, by looking at data taken from individual seminars and looking at the meta data of the collective set.

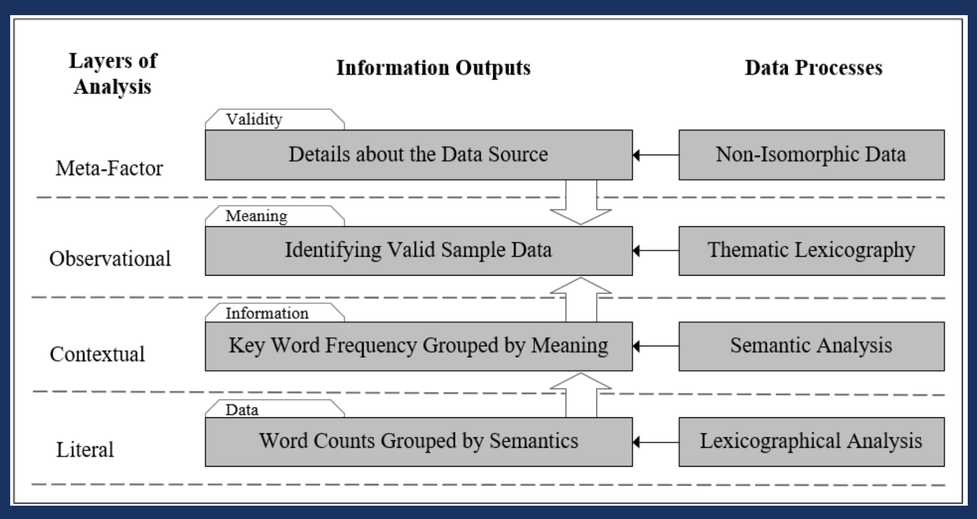

Figure 2 – A systematic, multi-modal model of interpretive content analysis for application of deep learning for complex qualitative data sets.

Problem Analysis and Formative Challenges

Interpretive Science
Grounded Theory
Qualitative Analysis

Automated Interpretation
Content Analysis

Empirical Science
Machine Learning
Quantitative Analysis

A. Limitations of Automated Qualitative Analysis

- Expensive to initially gather data
- Processing and analysis is highly subjective
- Hard to replicate, hard to control

Validation

B. Limitations of Procedural Quantitative Analysis

Verification

- Abstract and prone to omit non-isomorphic data
- Predisposed to unobserved biases and
- Can lacks semantic and contextual understanding

New Methods of Content Analysis and Experimental Findings

1. Literal Data Modelling

Lexicographical analysis and linear regression are applied to high-level data from the sample (such as seminar titles) and used to identify localised patterns within the data set.

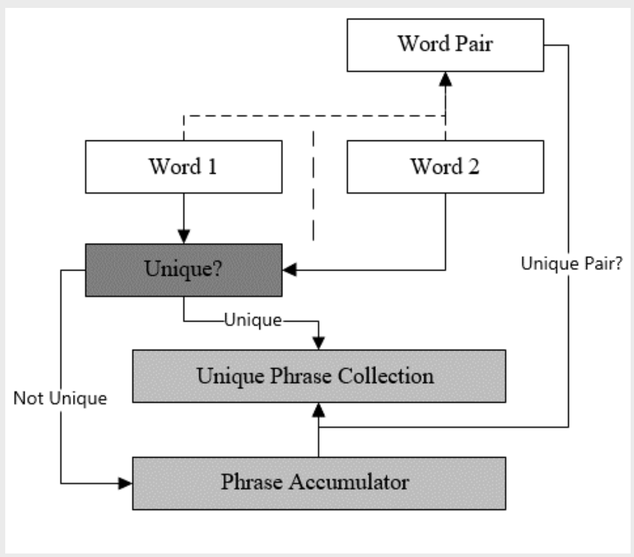

Figure 3 – A word-pairing algorithm combined with natural language processing is used to model key trends and factors at a high-level from the data set.

Tip: Lexicography models are not best developed in isolation.
A representative sample of professional experts or should help to train dictionary models for a given specialised corpus.

2. Observational Data Modelling

Word pairs are semantically grouped and modelled to find aggregate trends across the samples using recursive analysis.

	2017				2018				2019		
Ch	Rank	⁄	Concat	Ch	Rank	⁄	Concat	Ch	Rank	⁄	Concat
-12	3	19	Animation	-2	1	23	Reflection	-2	1	14	Your
-7	3	19	Indie	0	2	15	Player	0	1	14	Reflection
-7	9	8	Generation	-3	3	14	Your	-3	2	9	Art
-7	9	8	Procedural	-3	4	13	Story	-3	3	7	Creating
-6	10	7	Day	-4	5	12	Making	-3	3	7	Building
-6	10	7	No	-7	5	12	Art	-7	3	7	Bootcamp
-6	10	7	UX	-1	6	11	Creating	-1	4	6	Making
-5	2	21	Your	-9	6	11	Animation	-9	4	6	Play
-5	8	10	Character	-5	7	10	Level	-5	4	6	World
-5	11	6	Direction	2	7	10	Day	2	4	6	Player
-5	11	6	Our	-3	7	10	AI	-3	4	6	AI
-5	11	6	Programmer	-9	7	10	Building	-9	5	5	Your Game
-5	11	6	Social	-2	7	10	Development	-2	5	5	Level

Figure 4 – A recursive model is used to identify initial trends over aggregate data.

Tip: Compile a lexicography filtering out stop words and then compile one which retains them, analysing both.
Stop words lack semantic meaning when analysed in isolation, but add details to word pairs. E.g. 'Your game' is prescriptive, but 'our game' is reflective.

3. Contextual Language Modelling

Deep learning is applied to the full content of each qualitative data set to model key topics to form a corpus-based language model.

Tip: Dictation programs were found to be up to 37% more accurate than auto- captioning libraries.
'Big Data' on the cloud has allowed language process of timbre, accents, and intonation from speaker(s) to improve analysis and recognition.

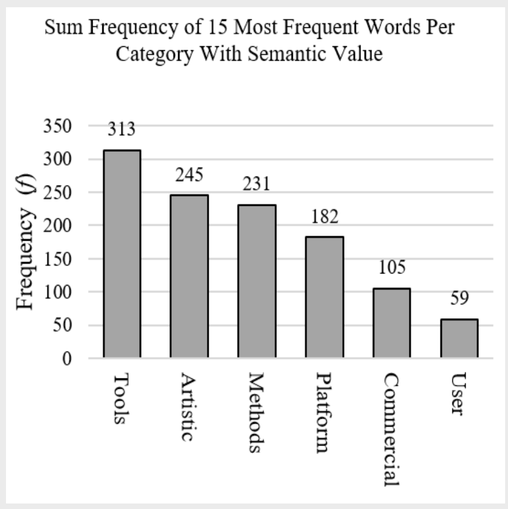

Figure 5 – A corpus-based semantic model is trained to recognise aggregated trends.

4. Meta-factor Data Source Modelling

Observations data-mined from the data source offers rich details to facilitate multiple-linear regression to identify trends, while maintaining power of the experiment as the sample size is further regressed and shrinks.

Figure 6 – The corpus-based semantic model is cross-referenced against meta-data from the source to find deeper trends from the aggregate data.

Tip: There is always non-isomorphic data which can't be anticipated or which will be tainted by observed or unobserved biases.

RESEARCHER: MR BRIAN WILLIAM PACKER
SUPERVISOR: PROFESSOR SIMEON KEATES

UNIVERSITY OF CHICHESTER