

IPS Workshop 2020



AIR CANADA



27 August 2020



The team

Université 
de Montréal



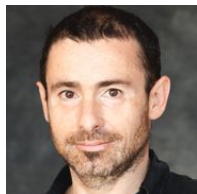
**Ilan
Elbaz**



**Vincent
Letard**



**David
Kletz**



**Philippe
Langlais**



**Frédéric
Piedboeuf**



**Lucas
Pagès**



**Fabrizio
Gotti**



**David
Alfonso**



**Tianjian
Gao**



Keith Dugas

 **AIR CANADA**



Nicholas Popovic

 **AIR CANADA**



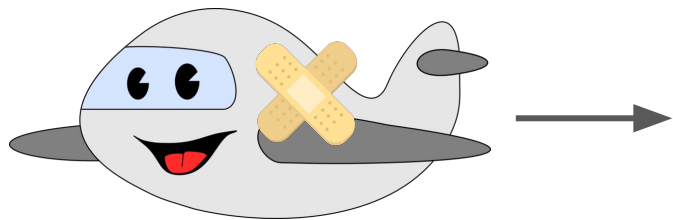
Tina - Yang Zhou



Helen Santos



Finding recurring aircraft defects



Defect report

**FLOOR LEVEL LIGHTING STRIP BY
17D MISSING**

Air Transport Association (ATA) code
classification: **33-50**
i.e. “emergency lighting, exit sign:aisle
inoperative”

How do we find defects that recur?

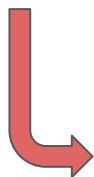
The corpus

Type	Description	Timestamp	Chapter	Section	MEL	Resolution
C	C/M POS 104 NEEDS RELAMPING FOR "HOT PLATE ON".	2018-01-04 02:13:00	25	0		RELAMPED.
E	GALLEY, WALL PANEL, LAMINATES, SCRATCHED, AT LOCATION:AFT, AT POSITION:FELL OFF THE WALL	2019-01-17 15:03:00	25	30		REPAIRED OK FOR SERVICE.
L	ON MORNING POWER UP HMU ADVISORY MSG APPEARED.	2019-12-28 13:13:00	46	0	491753	TRIAGE

- **460k** defect reports (2018-2019).
- **48** fields of data.

The corpus - Who inputs this data ?

Type	Description	Timestamp	Chapter	Section	MEL	Resolution
C	C/M POS 104 NEEDS RELAMPING FOR "HOT PLATE ON".	2018-01-04 02:13:00	25	0		RELAMPED.
E	GALLEY, WALL PANEL, LAMINATES, SCRATCHED, AT LOCATION:AFT, AT POSITION:FELL OFF THE WALL	2019-01-17 15:03:00	25	30		REPAIRED OK FOR SERVICE.
L	ON MORNING POWER UP HMU ADVISORY MSG APPEARED.	2019-12-28 13:13:00	46	0	491753	TRIAGE



- **C** : Cabin crew.
- **E** : Cabin crew using a form-like app.
- **L** : Cockpit crew and ground crew technicians.
- **MEL** : defect specialists.

The corpus - What's with Chapter and Section ?

Type	Description	Timestamp	Chapter	Section	MEL	Resolution
C	C/M POS 104 NEEDS RELAMPING FOR "HOT PLATE ON".	2018-01-04 02:13:00	25	0		RELAMPED.
E	GALLEY, WALL PANEL, LAMINATES, SCRATCHED, AT LOCATION:AFT, AT POSITION:FELL OFF THE WALL	2019-01-17 15:03:00	25	30		REPAIRED OK FOR SERVICE.
L	ON MORNING POWER UP HMU ADVISORY MSG APPEARED.	2019-12-28 13:13:00	46	0	491753	TRIAGE

- **Chapter and Section** are **labels** in a **taxonomy** describing the location, equipment and nature of all possible defects.

The "0" (in Section), is sometimes used (by the crew) when classifying the **defect isn't evident**.

chapter	section	description
24	61	dc electrical load distr
25	0	equipment furnishings, equipment/furnishings, fixed baby bassinet:inoperative

Data curation

From the corpus, we obtained 3 datasets

- **Full**

- 380K train, 308K non-zero.
- Populated by people of various skills
- Least reliable.

- **Trax**

- 47K total, 34K non-zero.
- ATA clustering (human-) verified. [cluster of 3+ ATA occurrences]
- More reliable than full.

- **Reliable**

- 34k total.
- Made of E-type and ATA re-labeled (by MEL specialists).
- Most reliable ATA annotation

Normalization

- **Reliably sourced lists** to replace :
 - Domain-specific acronyms
 - Airport codes
- **Spell checker** (Levenshtein distance + character swap) to correct :
 - Frequent unknown words from the data (specialized vocabulary)
 - Words of the English dictionary
- **Acronym miner** (Dynamic Alignment btw. acronym & left context) :

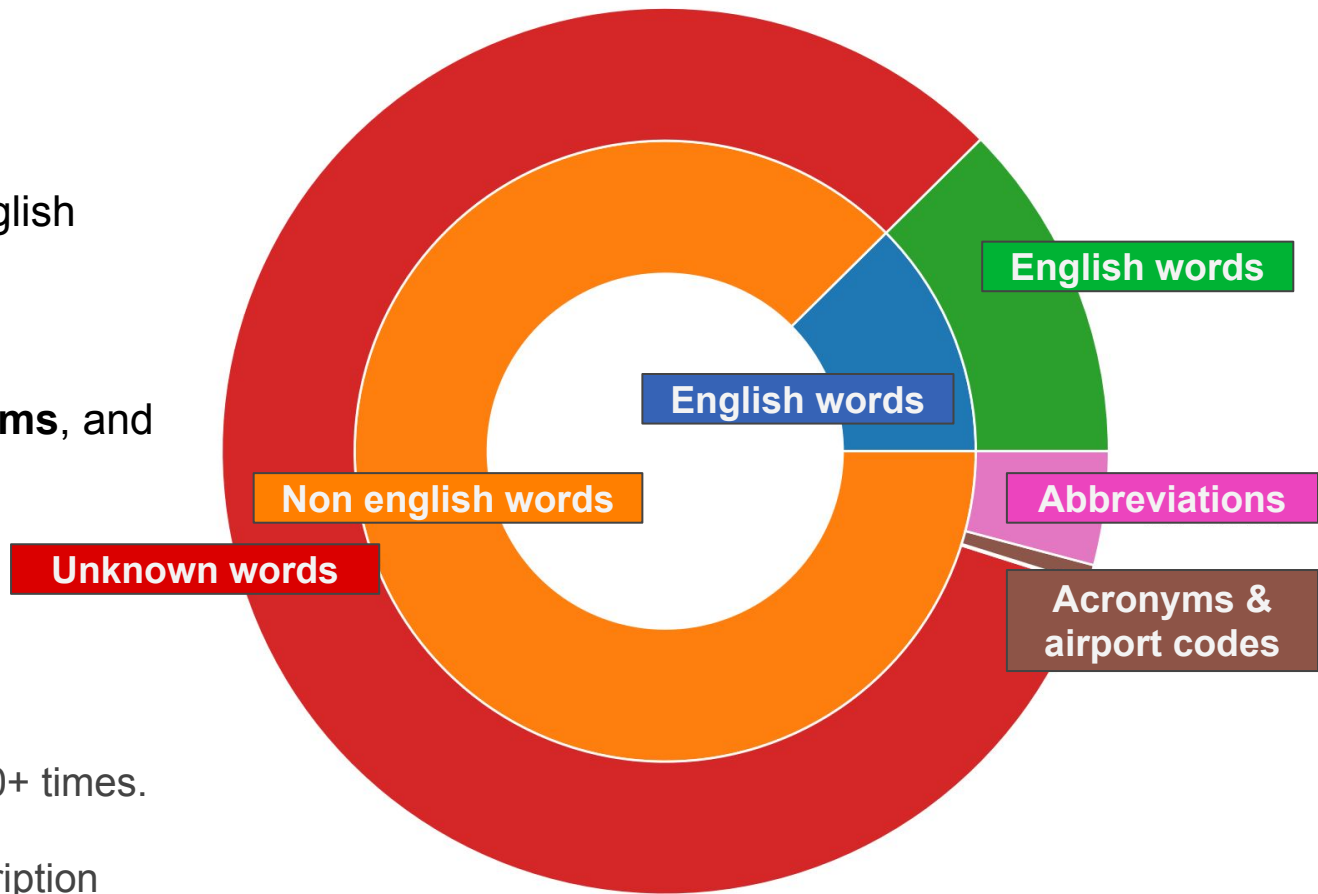
AUDIO/**V**IDEO **O**N **D**EMAND (AVOD)

ELECTRONIC **F**LIGHT **B**AG (EFB)

IN-**F**LIGHT **E**NTERTAINMENT(IFE)

Corpus Statistics

- **Vocabulary** size : **65k**
(64% hapax).
- **12%** **vocabulary** in English
lexicon.
- **5%** of the unk. voc. are
**abbreviations, acronyms, and
airport codes.**
- **83%** contains mostly
numbers, typos, etc.



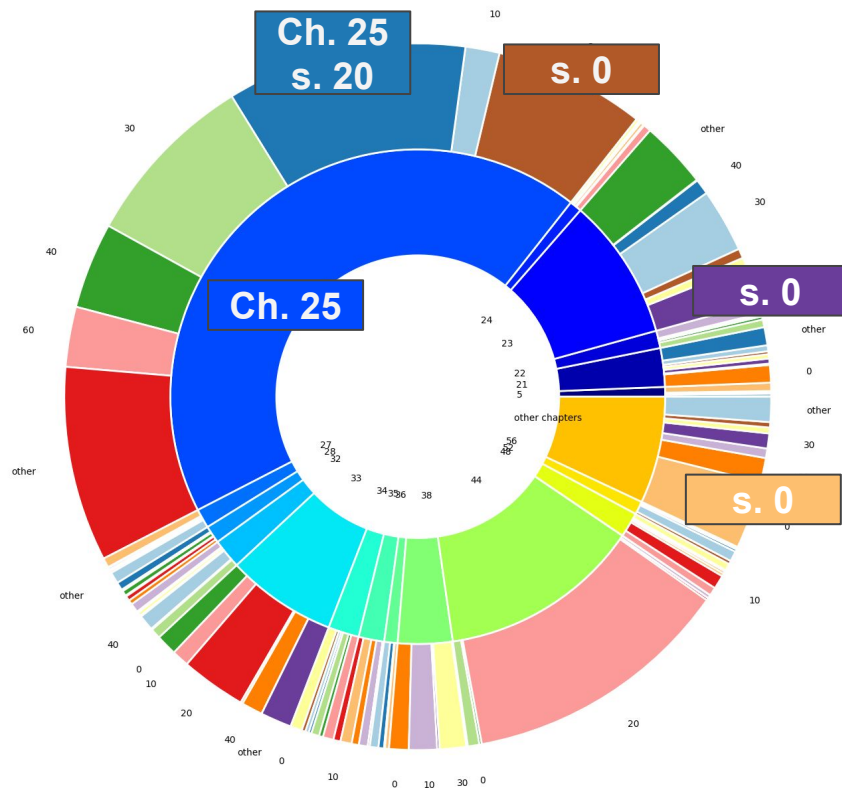
- Only 9911 types occur 10+ times.
- 26 defects have no description

Corpus Statistics

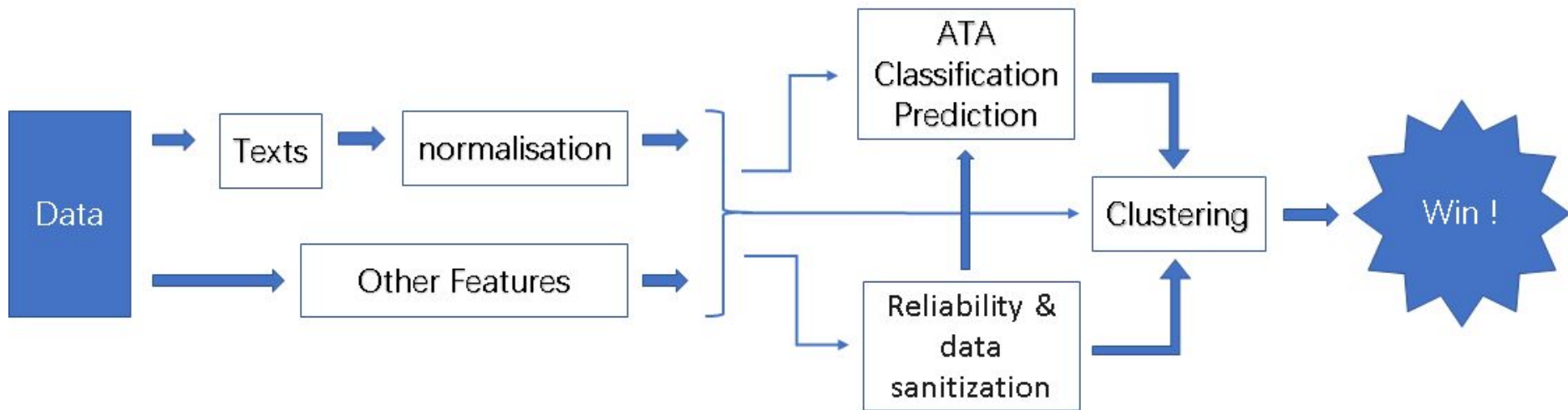
Very unbalanced ATA classifications:

- **43%** of the defects are classified in Chapter 25.
- **19%** of Sections have a value of 0.
- **11%** of Chapter-Sections pairs are classified as 25-20.

Pie chart of Chapter and Section distribution



The process



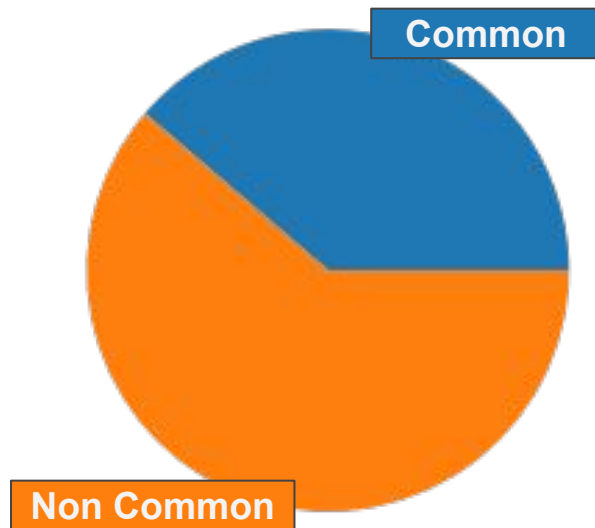
Classification - Issues with deep learning

- **GRU** (11.3) and **BERT** (18.4) could only learn the majority class.

Possibly due to the skewed ATA distribution.

- Weighting the samples did not help.
- Oversampling/undersampling awkward due to the skewness

Distribution of labels for the full dataset
(threshold=5%)



Classification - Why does TF-IDF work so well ?

class 11-32 : placard, placards, placcard, belongs, theres, damage, sticking, sure

class 21-20 : recirculation, fans, fan, gasper, recirc, installed, smell, present, recir

class 21-30 : auto, alt, outflow, cabin, tcn, pressure, indicator, rate, altitude, auto2

class 21-40 : heating, heater, heaters, heat, ovht, cargo, iii, duct, forward, vent

class 21-50 : pack, cooling, conditioning, deflector, ball, exhaust, packs, fcvs, bypass

class 21-60 : temp, zone, compt, modulating, overboard, trim, control, cabin, temperature

Classification - Trying to quantify the noise

Train	Test			
		Full	Trax	Reliable
	Full (max_voc=20k)	60.1	75.7	65.9
	Trax	22.5	81.8	37.9
	Reliable	27.0	50.8	97.9

Scores computed with a support vector classifier using token unigrams and bigrams.

Classification - Conclusion

- The task is **simple**, but the noise in the data makes it very **hard to classify**.
- High proportion of **specialized vocabulary**: pre-trained methods not applicable.
- Very imbalanced classes

F1 scores summary :

Dataset \ Algorithm	Dummy	BERT	Glove	Best SVC
Full	6.7	-	50.2	60.1
Trax	6.7	-	-	81.8
Reliable	36.2	18.4	-	97.9

Clustering

Clustering directly performed very poorly :

- The usual features :
 - Tf-Idf vectors, word embeddings
 - Distance in days (recurrence period mean length: 6.6 days, std: 9.7 days)
 - Equality of Chapter-Section
 - Mentions of one defect by another
- Balancing the nb. of the clusters with their average size is quite tricky
- Results were near 0, must investigate further
 - Either the reference is wrong/inconsistent
 - Or the algorithms used are way off

Conclusions

- A very **rich and complex** dataset requiring a lot of effort.
- **Many paths** (and dead-ends...) to achieving the same goal.
- **Classification** vastly outperforms direct clustering for now.
- **Clustering** via classification is the most promising approach.
- The task and the data **deserve** much **more work**.

ANNEX

Clustering

Original system clusters: recurrence length mean: 6.6 days, std: 9.7 days

