# "Understand this for me"

Software for Concept Analysis
David Upton
(Stirling Reid Limited)

# Operational analysis example

- In the Second World War, analysts proved that the RAF were not dropping bombs accurately.

- They did this by measuring distances on reconnaissance photos, between crater and target, this giving a precise measure of accuracy.

- As a result, the RAF introduced better systems and became able to drop bombs very much more accurately.

# Words.

- Words are the main tool of high level analysis
- But they can be misleading or ambiguous
- Meanings are given in dictionaries…
- And in official definitions (eg NATO glossary)
- But these work best for concrete words (**afterburning2 /postcombustion2:** The process of fuel injection and combustion in the exhaust jet of a turbojet engine (aft or to the rear of the turbine).01 Feb 1973
- Abstract words are more difficult

  "What then is time? If no one asks me, I know what it is. If I wish to explain it to him who asks, I do not know". (Saint Augustine)
- Even when a definition exists, do I use it?

# Language Games

- Concept developed by Ludwig Wittgenstein.
- 'Game' not used in a derogatory or trivialising sense: a 'game' is any process bound by rules.
- For instance poker and bridge: both involve people sitting at table moving standardised cards.
- The difference is in the rules: I would soon find this out if I tried to play poker at a bridge table.
- A language game is using words according to a set of (usually poorly defined) rules or assumptions.
- We all do this, all the time.

# Wittgenstein: Language Games

- My wife asks me to paint a door.

- I paint a picture of the door on canvas.

- She remonstrates with me: why did you not cover the door itself with house paint?

- I reply that the word 'paint' means both activities: why did she not say what she meant?

- It did not occur to her that her use of words was ambiguous – nor to me that my understanding could be wrong.

- We were playing different "language games"

# Ranges of meaning: 'affordability"

- Your customer asks you to find an 'affordable solution'. What does your customer mean?

- What may be 'affordable' to Bill Gates may not be affordable to me. (Affordable also implies 'acceptable', not just how much money you have.)

- Also, do you mean affordable in terms of money? Or time? Or resources?

- Or in terms of reputation damage? Or loss of personnel/ assets?

# Language games as group indicators

- Soldiers and sailors (or politicians and academics, civil servants, businesspeople…) use different words for similar things.

- You can tell one from the other just by their use of language.

- Partly this is a tribal badge, partly a means of coping with a different shared environment and activity.

- All are playing 'language games'.

# Operationalising this idea

- We are trying to find ways in which a "basic concept" is being skewed when it is used in a single text

- Just like a message being sent over a 'noisy channel'.

- Analysing the noise will show us the writer's language game

- Once we know that, we can understand their meaning better, and translate it into our own 'game'.

# Spies and cyphers

- During WW2, many secret agent messages sent to the UK were initially undecipherable.

- Experts could tell whether the problem was due to poor transmission ('morse mutilation') or mistakes in coding.

- Can we do the same with 'noisy channels' in texts?

# Noisy channel examples (1)

- You receive 'yhe' instead of 'the':
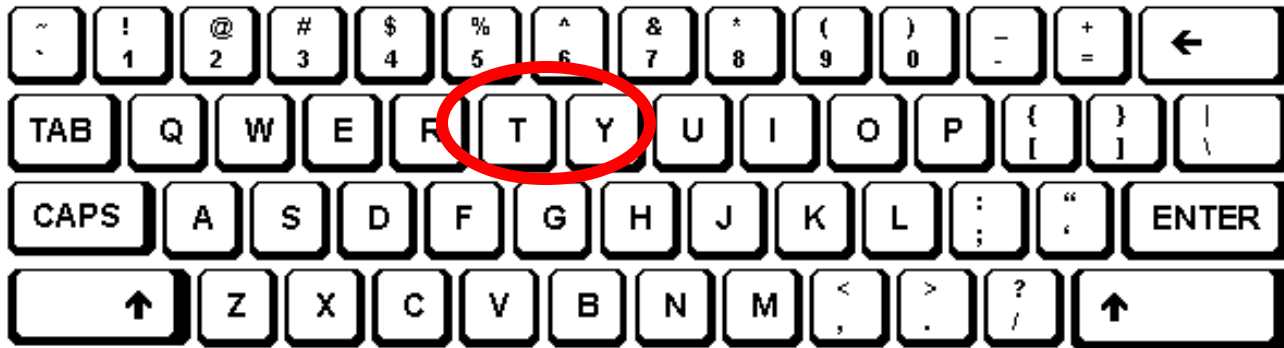


- Error explained by 'y' key being next to 't' on standard keyboard.
- Deduction: channel involved QWERTY keyboard.

# Noisy channel examples (2)

- You receive 'ind' instead of 'and'

- i = ..

- a = .-

- Error explained by one Morse code 'digit' being wrong.

- Deduction: channel involved morse code

# We need something to compare

- To examine the channel, we need some 'original idea' to compare with the text.
- We could learn a lot from comparing completely dissimilar texts – (eg Trattoria menu vs Prime Minister's speech)
- But the lessons would be very general
  - This is in Italian, that in English
  - This is short, that is long
  - This is a list, that is coherent prose
  - This has prices, that does not

# Selecting a corpus

- We need a 'corpus' of similar texts
- One approach is to choose documents 'by hand'.
- One approach is to define what exact criteria we want our selections to meet. This is not easy.
- Another is to pick out suitable examples, then find what they have in common….
- …and use this to develop a 'signature', against which to match new entrants.
- Surprisingly rapid means of document triage

# Dictionaries

- Lists of words and patterns that suggest certain 'language games'.

- These help us to know what to measure.

- For instance, what words show:
  - Money-biased game? (cost, expenditure, figures)
  - Reputation-biased game? (image, appear, media..0
  - Security-biased game? (secure, survivable, safety..)
  - Also need to disambiguate ( "bank"? River or financial? Cabinet "minister" vs Protestant "minister"?)

# Just a matter of counting

- Natural Language Processing (NLP) techniques developed to count word frequencies, patterns, etc.

- Faster and more accurate than humans.

- Surprisingly good AI – can mimic human judgements quite well.

# Results – an example

- You are asked to write a brief for a customer on 'affordable' options.
- You need to compare how the customer understands 'affordable' with your own concept of the word
- You look at how the customer's 'language game' compares with your own. What words / patterns do each use?
- You select relevant documents by the customer, and by your own organisation.
- Then you compare these from different viewpoints, eg:
  - Reputation?
  - Security?
  - Value for money?

# Summary

- A system to measure the 'meanings' of words, as they are used by different writers
- Avoid misunderstanding
- Address requirements more precisely
- Assess opposing views more clearly
- Better document triage.
- David@simulation.cc
- (+44) 7957 597904