Natural Language Generation

Dr. Nava Tintarev
Assistant Professor, TU Delft

Credits: These slides are modified from Prof. Ehud Reiter (Aberdeen Uni and Arria/Data2text)

TU Delft

Common Machine Learning techniques

- Common ML techniques
  - Discriminative: SVM, MaxEnt/Log. Reg.
  - Generative: NB
  - **Discriminative v. Generative**

Naive Bayes vs. Maxent Models

- Naive Bayes models multi-count correlated evidence
    - Each feature is multiplied in, even when you have multiple features telling you the same thing

- Maximum Entropy models (pretty much) solve this problem
    - This is done by weighting features so that model expectations match the observed (empirical) expectations

TU Delft

Generative (joint models)

$P(c,d)$

- We have some data {(*d*, *c*)} of paired observations *d* and hidden classes *c*.
- Joint (generative) models place probabilities over both observed data and the hidden stuff (generate the observed data from hidden stuff):
  - All the classic Statistical NLP models:
    - *n*-gram models, Naive Bayes classifiers, hidden Markov models, probabilistic context-free grammars, IBM machine translation alignment models

TU Delft

Generative (joint)

$P(c,d)$

- Easy to train: just count
- Language modeling: probability of observed forms
- More robust
  - Small training sets
  - Label noise
- Full advantage of probabilistic methods

TU Delft

Discriminative (conditional) Models

P (c | d)

- Discriminative (conditional) models take the data as given, and put a probability over hidden structure given the data:
    - Logistic regression, conditional loglinear or maximum entropy models, conditional random fields
    - Also, SVMs, (averaged) perceptron, etc. are discriminative classifiers (but not directly probabilistic)

TU Delft

Discriminative (conditional)

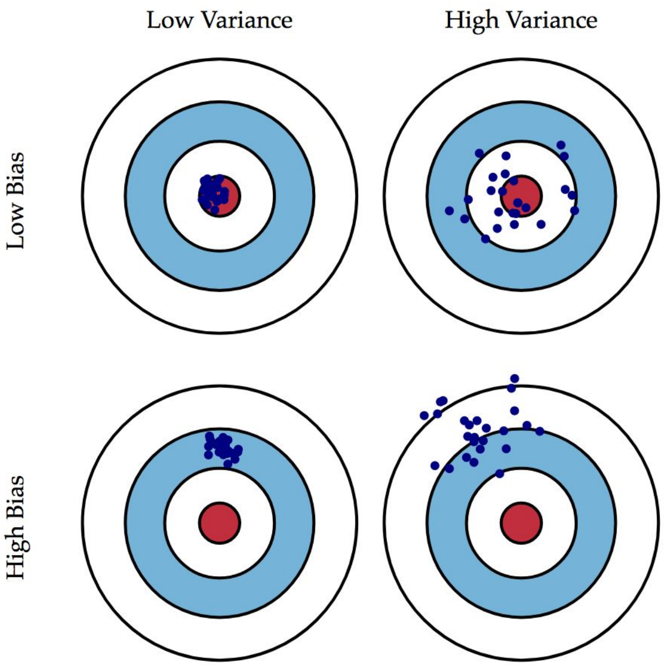
P (c | d)

- Give high **accuracy** performance
- They make it easy to incorporate **lots of linguistically important features**
- They allow automatic building of **language independent,** retargetable NLP modules

TUDelft

Practical issues

Bias versus variance trade-off



Error due to Bias: The error due to bias is taken as the difference between the expected (or average) prediction of our model and the correct value which we are trying to predict.
Think over-optimizing precision.

Error due to Variance: The error due to variance is taken as the variability of a model prediction for a given data point.
Think over-optimizing recall.

Very little data?

- Use Naïve Bayes
  - Naïve Bayes is a "high-bias" algorithm (Ng and Jordan 2002 NIPS)
- Get more labeled data
  - Find clever ways to get humans to label data for you
- Try semi-supervised training methods:
  - Bootstrapping, Expectation maximization over unlabeled documents, …

TU Delft

A reasonable amount of data?

- Perfect for all the clever classifiers
  - SVM
  - Logistic Regression
- You can even use user-interpretable decision trees
  - Users like to hack
  - Management likes quick fixes

11

A huge amount of data?

- Can achieve high accuracy!
- At a cost:
  - SVMs (train time) or kNN (test time) can be too slow
  - Regularized logistic regression can be somewhat better
- So Naïve Bayes can come back into its own again!

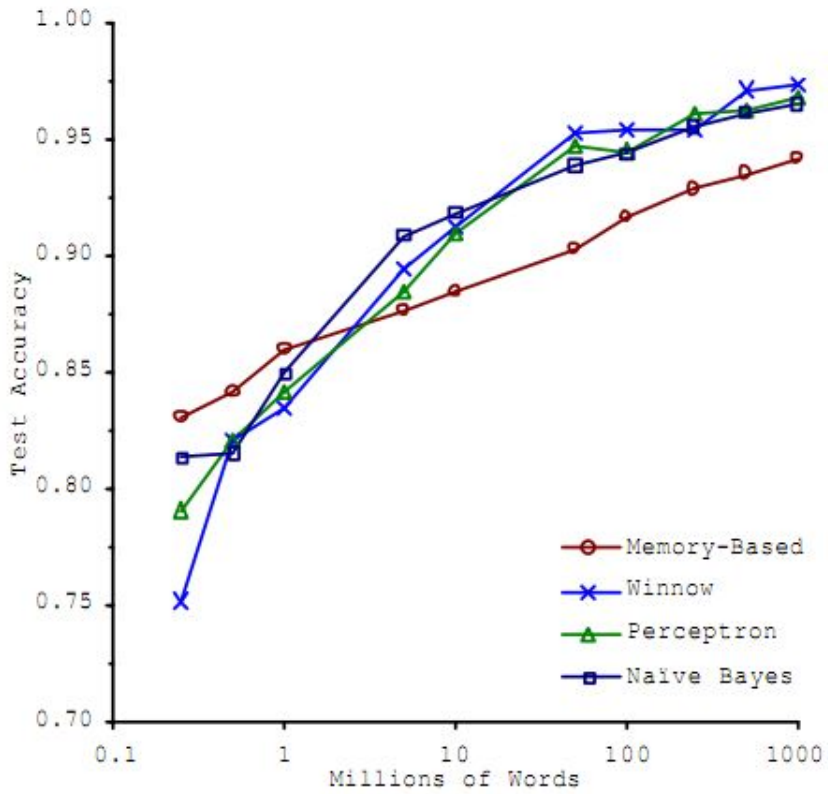
How much data?

- **Factor of the number of classes**: There must be x independent examples for each class, where x could be tens, hundreds, or thousands (e.g. 5, 50, 500, 5000).
- **Factor of the number of input features**: There must be x% more examples than there are input features, where x could be tens (e.g. 10).
- **Factor of the number of model parameters**: There must be x independent examples for each parameter in the model, where x could be tens (e.g. 10).

TU Delft

Accuracy as a function of data size

With enough data, classifier may not matter
...but there is a risk over-fitting!



Brill and Banko on spelling correction

TU Delft

Real-world systems generally combine:

A) Automatic classification AND
B) Manual review of uncertain/difficult/"new” cases

How to tweak performance

- Domain-specific features and weights: *very* important in real performance
- Sometimes need to collapse terms:
  - Part numbers, chemical formulas, …
  - But stemming generally doesn't help
- Upweighting: Counting a word as if it occurred twice:
  - title words (Cohen & Singer 1996)
  - first sentence of each paragraph (Murata, 1999)
  - In sentences that contain title words (Ko *et al,* 2002)

16

Questions?

Today

Natural language generation

- NLG pipeline
  - Document planning
  - Microplanning
  - Realisation
- Example systems
  - e.g., ScubaText

Natural Language Generation

COMFREAK@PIXABAY

What is NLG?

The sub-field of AI and CL …

that is concerned with the construction of …

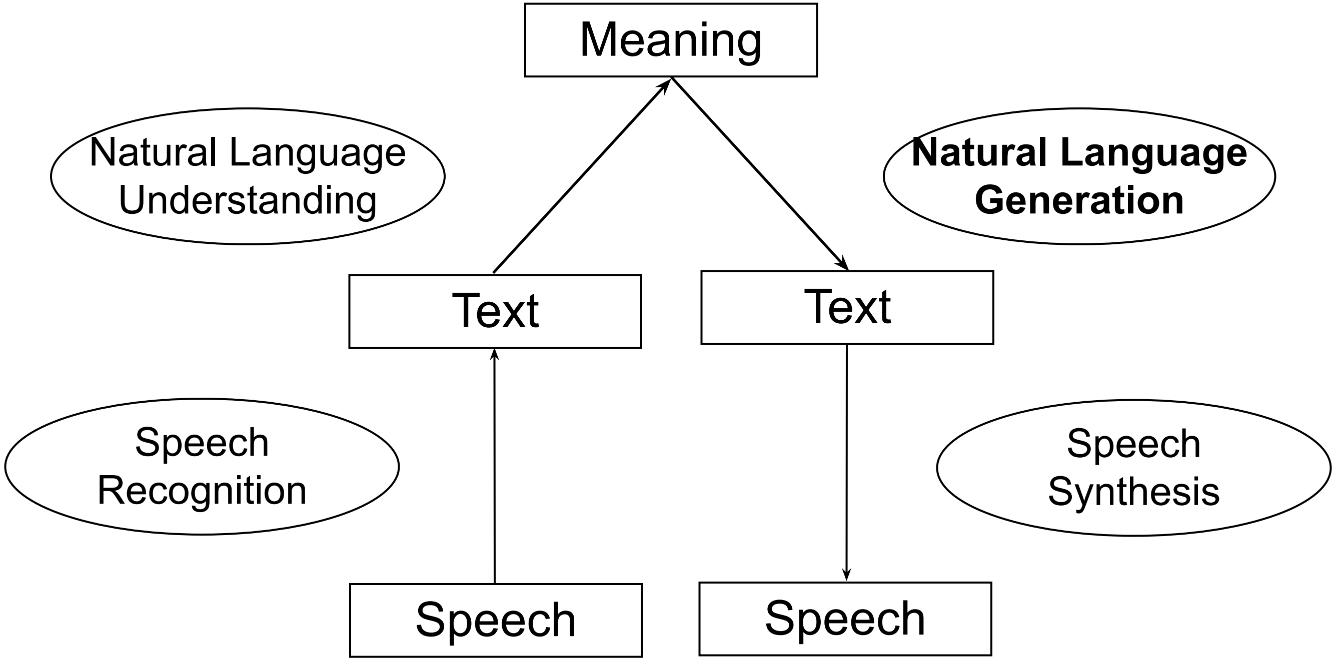
computer systems that can **produce understandable texts** in English or other human languages **from** some underlying **non-linguistic representation** of information.

What is NLG?

- NLG systems are computer systems which produces understandable and appropriate texts in English or other human languages
  - Input is data (raw, analysed)
  - Output is documents, reports, explanations, help messages, and other kinds of texts
- Requires
  - Knowledge of language
  - Knowledge of the domain

TU Delft

Language Technology



Data-to-text generation

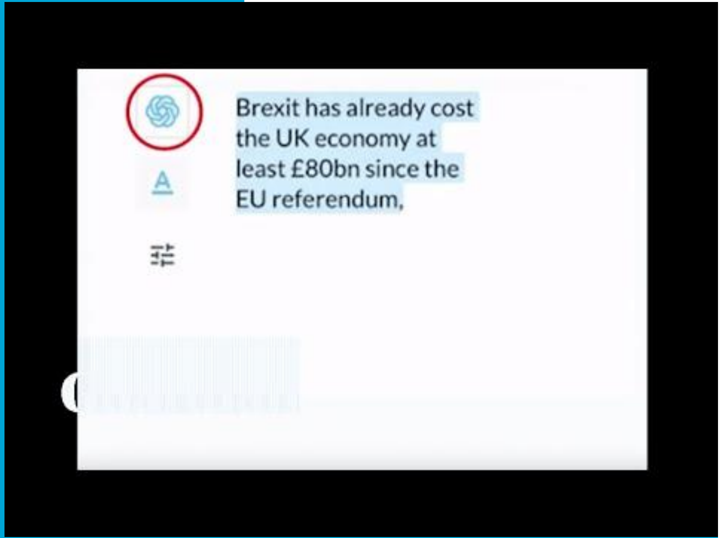
	Report for fans of Achilles '29	Report for fans of Dordrecht
Dutch	**Thoone velt Dordrecht: 2-1** Jop van Steen en Freek Thoone hebben ervoor gezorgd dat de uitploeg zonder punten achterbleef. In Groesbeek werd voor 1022 toeschouwers met 2-1 gewonnen van Dordrecht. De uitploeg kwam na 10 minuten uit het niets op een 0-1 voorsprong door een prachtige treffer van Janga. Jop van Steen schoot in de 48e minuut de dik verdiende gelijkmaker tegen de touwen. Thoone bracht na 88 minuten de winnende treffer op het scorebord: 2-1. Scheidsrechter Van den Kerkhof was genoodzaakt 3 gele kaarten te geven, aan Arnaud De Greef, Boy van de Beek en Josimar Lima.	**Het zit Dordrecht niet mee tegen Achilles '29: 2-1** De uitploeg leed een zure nederlaag uit tegen de ploeg van manager Eric Meijers. Dordrecht verloor na een hoopvol begin met 2-1 van Achilles '29. Aanvaller Janga zette de ploeg van manager Gérard de Nooijer op een 0-1. Achilles '29 kwam door twee gelukkige treffers van Van Steen en Freek Thoone op een 2-1 voorsprong. Er werden 3 gele kaarten uitgedeeld: aan de zijde van Dordrecht voor Arnaud de Greef en Josimar Lima en aan de zijde van de thuisploeg voor Boy van de Beek.

- Soccer reports
- Virtual 'newspapers' from sensor data
- Wildlife tracking, feedback to citizen scientists
- Weather and financial reports
- Patient information
- Information about cultural artifacts….

Does NLG include text-to-text?

- Text-to-text **input** less varied
- Text may be one of many kinds of input to an NLG system.

Text-to-text generation



GPT-2
https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction
https://openai.com/blog/better-language-models/

- Machine translation
- Fusion and summarization of related sentences
- Simplification of complex texts (e.g. readability level)
- Spelling/grammar/text correction
- Peer review of scientific papers
- Automatic generation of questions (e.g., for education)
- Fake research papers:
  - https://pdos.csail.mit.edu/archive/scigen/

Text simplification

Advaith Siddharthan. "Syntactic Simplification and Text Cohesion". In Research on Language and Computation, Volume 4, Issue 1, Jun 2006, Pages 77–109, Springer Science, the Netherlands.

- **Syntactic Simplification:** process that reduces the syntactic complexity of a text while preserving its meaning and information content.
- **Lexical Simplification:** process that reduces the lexical complexity of a text while preserving its meaning and information content.

Example syntactic simplification:

- *Also contributing to the firmness in copper, the analyst noted, was a report by Chicago purchasing agents, which precedes the full purchasing agents report that is due out today and gives an indication of what the full report might hold.*
- *Also contributing to the firmness in copper, the analyst noted, was a report by Chicago purchasing agents. The Chicago report precedes the full purchasing agents report. The Chicago report gives an indication of what the full report might hold. The full report is due out today.*

NB: Not shorter, but more cohesive.

Syntactic Simplification (Siddharthan)

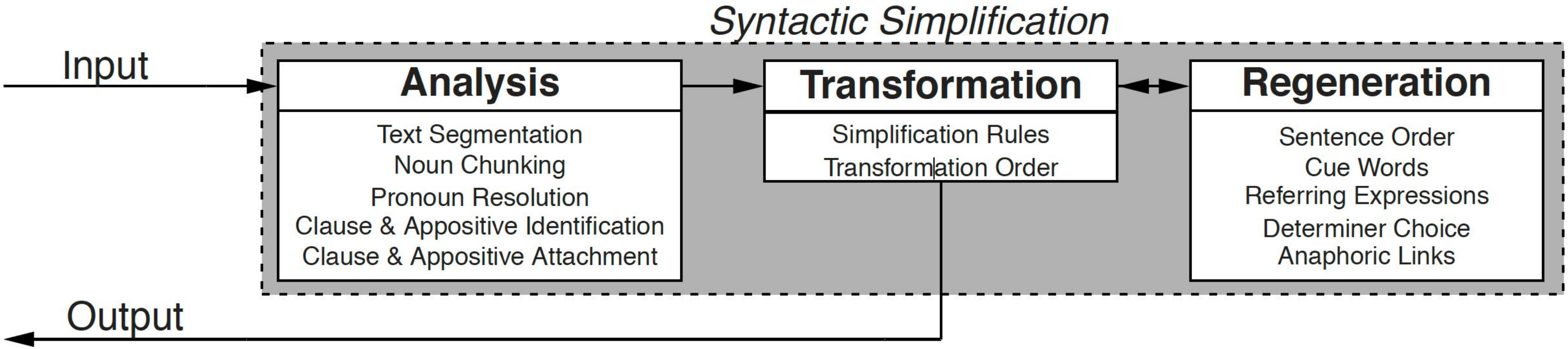


Figure 2.1. An architecture for a text simplification system

Gricean maxims

- Maxim of Quantity: Be exactly as informative as is required.
- Maxim of Quality: Try to make your contribution one that is true.
- Maxim of Relevance: Be relevant.
- Maxim of Manner: Be perspicuous (~clear)

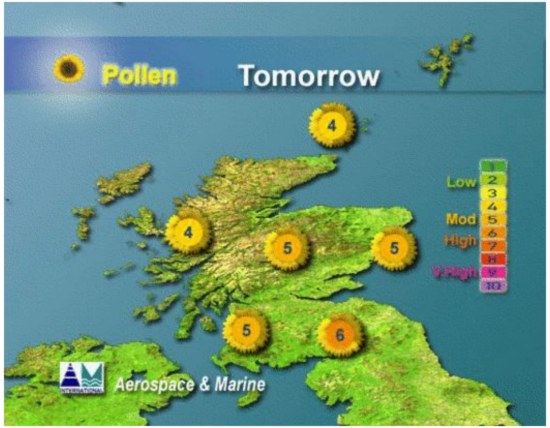
Some benefits of NLG

Recall previous lectures

- Weather forecast by NLG system preferred to experts!
- More consistent, better word choice
- NLG in clinical setting helped make better decisions than graphics!

TU Delft

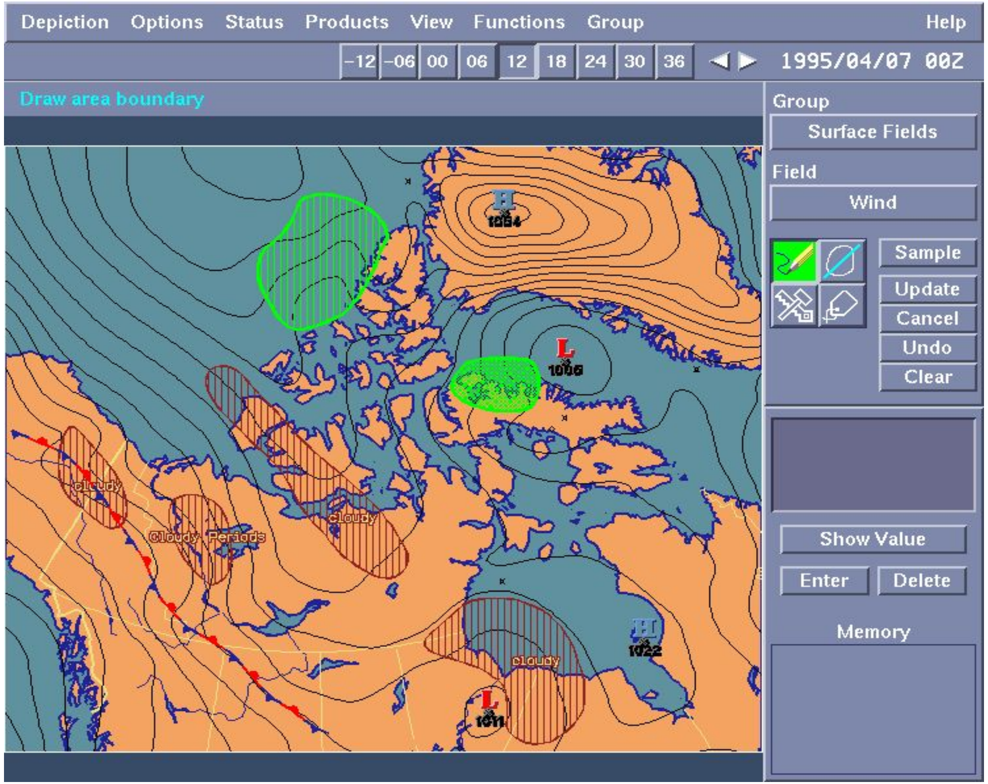
Simple ex: Pollen forecasts



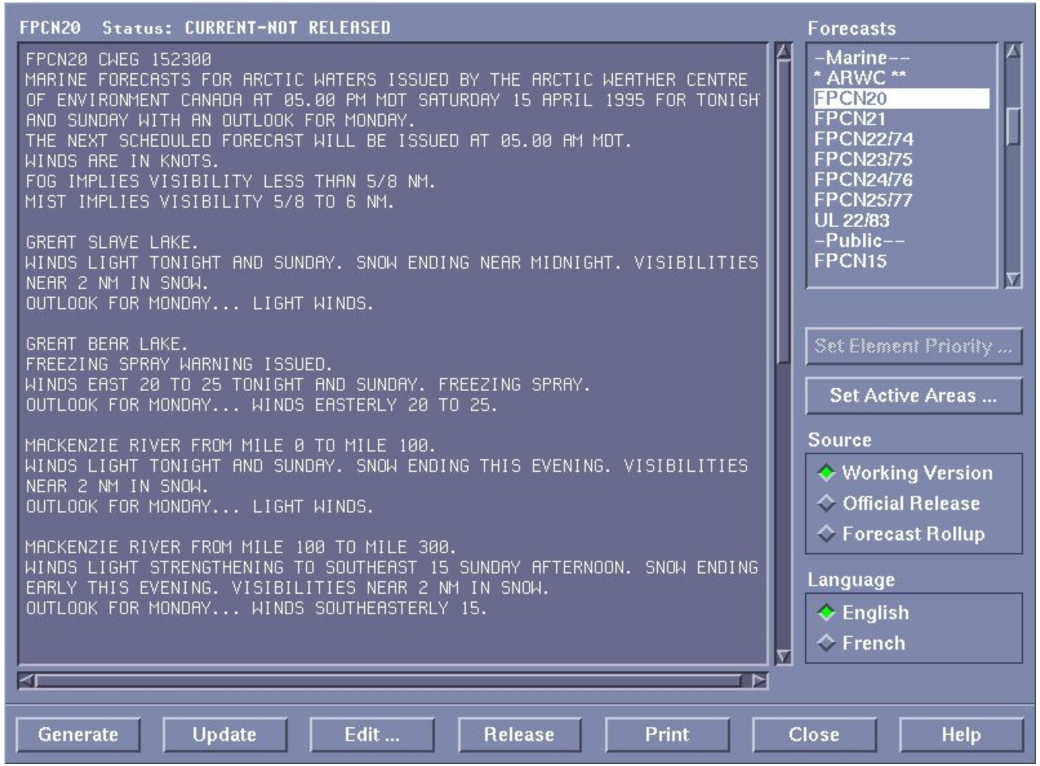

```
Grass pollen levels for Tuesday have
decreased from the high levels of yesterday
with values of around 4 to 5 across most
parts of the country. However, in South
Eastern areas, pollen levels will be high
with values of 6.
```

TU Delft

Medium ex: marine forecasts



FoG: Output


FPCN20 Status: CURRENT-NOT RELEASED
FPCN20 CWEG 152300
MARINE FORECASTS FOR ARCTIC WATERS ISSUED BY THE ARCTIC WEATHER CENTRE
OF ENVIRONMENT CANADA AT 05.00 PM MDT SATURDAY 15 APRIL 1995 FOR TONIGH
AND SUNDAY WITH AN OUTLOOK FOR MONDAY.
THE NEXT SCHEDULED FORECAST WILL BE ISSUED AT 05.00 AM MDT.
WINDS ARE IN KNOTS.
FOG IMPLIES VISIBILITY LESS THAN 5/8 NM.
MIST IMPLIES VISIBILITY 5/8 TO 6 NM.
GREAT SLAVE LAKE.
WINDS LIGHT TONIGHT AND SUNDAY. SNOW ENDING NEAR MIDNIGHT. VISIBILITIES
NEAR 2 NM IN SNOW.
OUTLOOK FOR MONDAY... LIGHT WINDS.
GREAT BEAR LAKE.
FREEZING SPRAY WARNING ISSUED.
WINDS EAST 20 TO 25 TONIGHT AND SUNDAY. FREEZING SPRAY.
OUTLOOK FOR MONDAY... WINDS EASTERLY 20 TO 25.
MACKENZIE RIVER FROM MILE 0 TO MILE 100.
WINDS LIGHT TONIGHT AND SUNDAY. SNOW ENDING THIS EVENING. VISIBILITIES
NEAR 2 NM IN SNOW.
OUTLOOK FOR MONDAY... LIGHT WINDS.
MACKENZIE RIVER FROM MILE 100 TO MILE 300.
WINDS LIGHT STRENGTHENING TO SOUTHEAST 15 SUNDAY AFTERNOON. SNOW ENDING
EARLY THIS EVENING. VISIBILITIES NEAR 2 NM IN SNOW.
OUTLOOK FOR MONDAY... WINDS SOUTHEASTERLY 15.
Forecasts
-Marine--
* ARWC **
FPCN20
FPCN21
FPCN22/74
FPCN23/75
FPCN24/76
FPCN25/77
UL 22/83
-Public--
FPCN15
Set Element Priority ...
Set Active Areas ...
Source
Working Version
Official Release
Forecast Rollup
Language
English
French
Generate
Update
Edit ...
Release
Print
Close
Help

Complex example: road maintenance

- Forecasts for gritting and other winter road maintenance procedures
- Input is <u>15</u> parameters over space and time
  - Temperature, wind speed, rain, etc
  - Over thousands of points on a grid
  - Over 24 hours (20-min interval)

TU Delft

Points

Generated Text

Overview. Road surface temperatures will reach marginal levels on most routes from this evening until tomorrow morning.

Wind (mph). NW 10-20 gusts 30-35 for a time during the afternoon and evening in some southwestern places, veering NNW then backing NW and easing 5-10 tomorrow morning.

Weather. Light rain will affect all routes this afternoon, clearing by 17:00. Fog will affect some central and southern routes after midnight until early morning and light rain will return to all routes. Road surface temperatures will fall slowly during this afternoon until tonight, reaching marginal levels in some places above 200M by 17:00.

TUDelft

BabyTalk

- **Goal:** Summarise clinical data about premature babies in neonatal ICU
- **Input:** sensor data; records of actions/observations by medical staff
- **Output:** multi-para texts, summarise
  - BT45: 45 mins data, for doctors
  - BT-Nurse: 12 hrs data, for nurses
  - BT-Family: 24 hrs data, for parents

TU Delft

BT45 texts (extract)

Computer-generated text

- By 11:00 the baby had been hand-bagged a number of times causing 2 successive bradycardias. She was successfully re-intubated after 2 attempts. The baby was sucked out twice.

  At 11:02 FIO2 was raised to 79%.

Human corpus text

- At 1046 the baby is turned for re-intubation and re-intubation is complete by 1100 the baby being bagged with 60% oxygen between tubes. During the re-intubation there have been some significant bradycardias down to 60/min, but the sats have remained OK. The mean BP has varied between 23 and 56, but has now settled at 30. The central temperature has fallen to 36.1°C and the peripheral temperature to 33.7°C. The baby has needed up to 80% oxygen to keep the sats up.

TU Delft

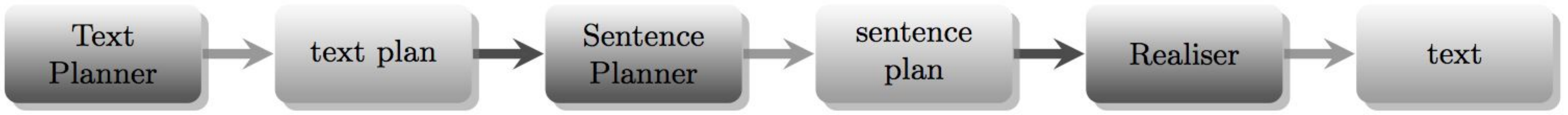
Other NLG projects

- NLG is not just for experts!

- Blogging birds: generate “blogs” from red kites based on location data
- Standup: help children with learning disabilities tell jokes
- Skillsum: give adults feedback on literacy/numeracy assessment
- How was school today…? Helping children talk about their day
- MinkApp: Motivating nature conservation volunteers

How do NLG Systems Work?

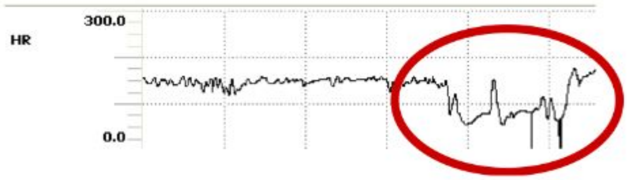
- Usually three stages
  - Not including data analysis
- Document planning *(content determination)*: decide on content and structure of text
- Microplanning: decide how to linguistically express text (which words, sentences, etc to use)
- Realisation: actually produce text, conforming to rules of grammar (e.g., referring expressions)



Questions?

Content determination

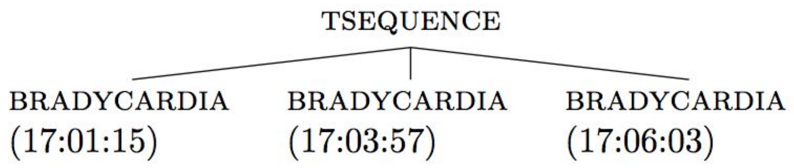
- Of the zillions of things I could say, which should I say?
  – Depends on what is important
  – Also depends on what is easy to say



(a) Content Determination

Text structuring

- E.g., order of presentation
- Applying discourse relations, e.g., contrasts or elaborations.
  - Hand-crafted structuring rules: Schemata
  - Rhetorical Structure Theory (RST)



(b) Text Structuring

Sentence aggregation

1. Sadio Mane scored for Southampton after 12 minutes and 22 seconds.
2. Sadio Mane scored for Southampton after 13 minutes and 46 seconds.
3. Sadio Mane scored for Southampton after 15 minutes and 18 seconds.
4. → Sadio Mane scored **three times** for Southampton in **less than three minutes**.

TU Delft

Lexicalisation

"To score a goal"

"To have a goal noted"

"To put the ball in the net"

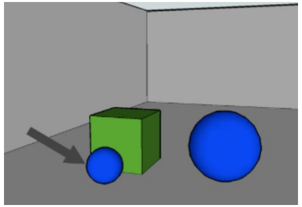
- Affective stance of the reader?
- Style constraints?
- Variation?

TUDelft

Referring Expression Generation

“The task of selecting words or phrases to identify domain entities”

- Referential form e.g., pronoun, proper name, (in)definite description
- Referential context
- Unique and not too long
  - “The small blue ball before the large green cube”/”the ball”.



REG/GRE algorithms

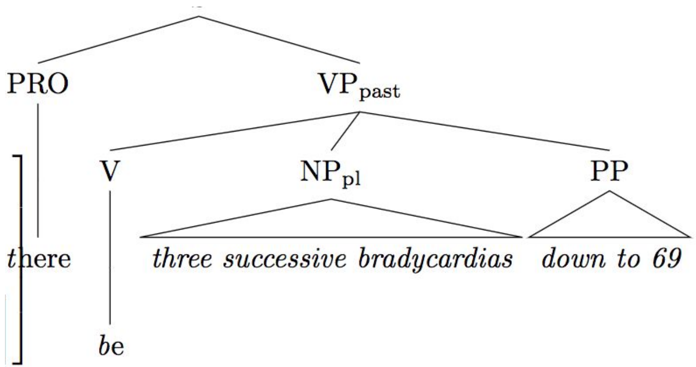
- **Full Brevity.** Exhaustive search, select shortest.
- **Greedy Heuristic.** Chose rule that excludes the most distractors at each step.
- **Incremental algorithm.** Based on domain-specific preference or cognitive salience.

TU Delft

Realisation

"There were three successive bradycardias down to 69."

- Human-crafted templates;
- Human-crafted grammar-based systems;
- Statistical approaches



(d) Realisation

Template-based realization

Hello Nava,

Thank you so much for the wonderful letter you wrote. Mrs Claus and I receive hundreds of letters each day. We look forward to sitting down in front of the fireplace in the evening to read them. You'll be excited to hear that the elves in my workshop are working very hard to be ready for Christmas!

I'm so pleased to hear that you have been so well behaved this year. I'm impressed that you have made time for your friends and family this year and not had too much to drink. I'm sure you have some wonderful plans together for Christmas too! Have fun!

I'm always pleased to see how kind you are and I'm proud of you. Rudolph and I are looking forward to our visit. Don't forget to leave him some carrots, it's a long journey!

A very merry Christmas to you and your family!

HO! HO! HO!

Santa

Yseop:

https://yseop.com/blog/letters-from-santa/

Realiser

- Just tell realiser verb, tense, whether negated, and it will figure out the rest
  - `(watch, future) -> will watch`
  - `(watch, past, negated) -> did not watch`

- Similarly automate other “obscure” encodings of information

Realiser

- Adjective ordering
  - <u>Big red </u>apple vs <u>Red big </u>apple
- Agreement and measurements
  - Three miles <u>is</u> a long way
  - Three children <u>are</u> hungry
- Bare infinitives and perception verbs
  - I see John <u>eat</u> an apple
  - I see John <u>thinks</u> a lot

Realiser - morphology

- Words have different forms
- Nouns have plural
  - Dog, dogs
- Verbs have base, present 3s, past, present participle, past participle
  - break, breaks, broke, breaking, broken
- Adjectives have comparative, superlative
  - Big, bigger, biggest

Realiser - morphology

Example: plural

- Usually add “s” (*dogs*)
  - But add “es” if base noun ends in certain letters (*boxes*, *guesses*)
  - Also change final “y” to “i” (*tries*)
- Many special cases
  - –*children* (vs *childs*), *people* (vs *persons*), etc

Realiser

- Calculates variants automatically
- (*dog*, plural) → *dogs*
- (*box*, plural) → *boxes*
- (*child*, plural) → *children*

- Also, punctuation, spacing etc

Realiser systems

- simplenlg – relatively limited functionality, but well documented, fast, easy to use, tested
- KPML – lots of functionality but poorly documented, buggy, slow
- openccg – somewhere in between
- ...

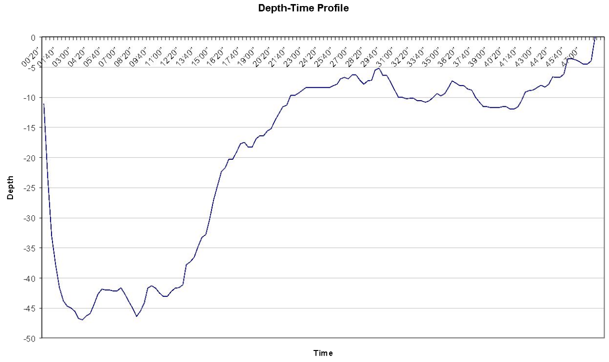
Scubatext example

- Demo system (Dr Sripada) for scuba divers
- Input is *dive computer data*
  - Depth-time profile of scuba dive
- Output is feedback to diver
  - Mistakes, what to do better next time
  - Encouragement of things done well

TU Delft

Scuba - input

Risky dive with some minor problems. Because your bottom time of 12 min exceeds no-stop limit by 4 min this dive is risky. But you performed the ascent well. Your buoyancy control in the bottom zone was poor as indicated by ‘saw tooth’ patterns.



Scuba: data analytics

- Look for trends and patterns in data
  - **Trends:** e.g., `depth increases fairly steadily over first 3 minutes`
  - **Patterns:** e.g., `sawtooth between 3 and 15 minutes`
- Will not further discuss here

Document Planning

- *Content selection*: Of the zillions of things I could say, which should I say?
  - Depends on what is important
  - Also depends on what is easy to say
- *Structure*: How should I organise this content as a text?
  - What order do I say things in?
  - Rhetorical structure?

Scuba: content

- Probably focus on patterns indicating dangerous activities
  – Most important thing to mention
- How much should we say about these?
  – Detail? Explanations?
- Should we say anything for safe dives?
  – Maybe just acknowledge them?
  – But encouragement also important

Scuba: structure

- Mention most dangerous thing first?
  - Or should we just order by time?
  - Start with overview?
- Linking words (cue phrases)
  - Also, but, because, …

TU Delft

Microplanning

- **Lexical/syntactic choice:** Which words and linguistic structures to use?
- **Aggregation:** How should information be distributed across sentences and paras
- **Reference:** How should the text refer to objects and entities?

SCUBA: microplanning

- **Lexical/syntactic choice:**
  - *Risky* vs *dangerous* vs *unwise* vs …
  - *Performed the ascent* vs *ascended* vs …
  - *12 min* vs *720 sec* vs *700 sec* vs *714.56 sec*
- **Aggregation:** 1 sentence or 2 sent?
  - "Because your bottom time of 12 min exceeds no-stop limit by 4 min this dive is risky, but you performed the ascent well."

TU Delft

Scuba: Microplanning

- **Aggregation** (continued)
  - Phrase merging
    - “Your first ascent was fine. Your second ascent was fine”   vs
    - “Your first and second ascents were fine.”
  - **Reference (appositive)**
    - Your ascent   vs
    - Your first ascent   vs
    - Your ascent from 33m at 3 min

Realisation

- **Grammars (linguistic)**: Form legal English sentences based on decisions made in previous stages
  - Obey sublanguage, genre constraints
- **Structure:** Form legal HTML, RTF, JSON, or whatever output format is desired

TU Delft

Scuba: Realisation

- **Simple linguistic processing**
  - Capitalise first word of sentence
  - Subject-verb agreement
    - Your first ascent <u>was</u> fine
    - Your first and second ascents <u>were</u> fine
- **Structure**
  - Inserting line breaks in text
  - Add HTML markups, eg, <P>

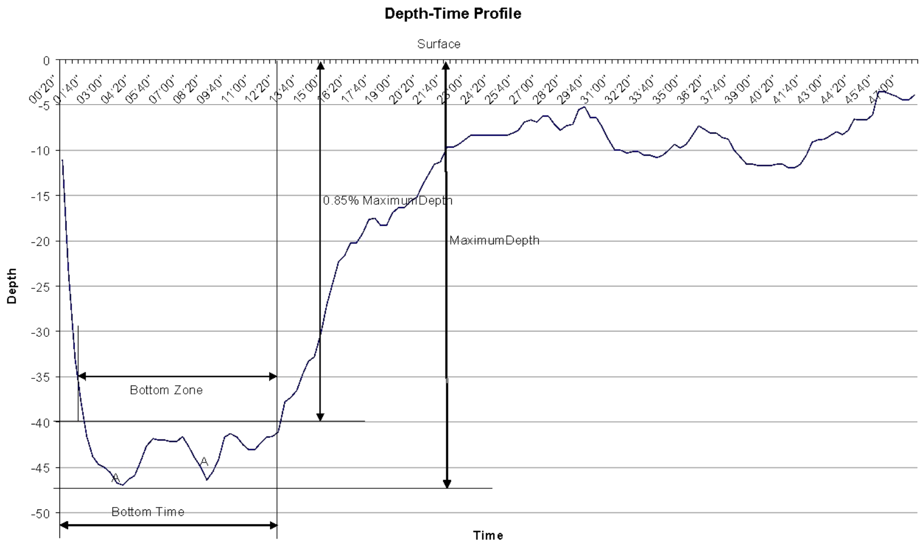
TU Delft

Multimodal NLG

- Speech output
- Text and visualisations
  - Produce separately, OR
  - Tight integration
    - E.g., text refers to graphic, OR
    - graphs has text annotations

TU Delft

Combined (Preferred)



Risky dive with some minor problems. Because your bottom time of 12.0min exceeds no-stop limit by 4.0min this dive is risky. But you performed the ascent well. Your buoyancy control in the bottom zone was poor as indicated by 'saw tooth' patterns marked 'A' on the depth-time profile.

Mitchell, M., Dodge, J., Goyal, A., Yamaguchi, K., Stratos, K., Han, X., Mensch, A., Berg, A., and Berg, T. L., Daume III, H. (2012). [Midge: Generating Image Descriptions From Computer Vision Detections](). *Proceedings of EACL 2012.*

TUDelft

Caption generation (NLG+Vision)

- Computer vision
  - Object detection
  - Holistic scene analysis
  - Dense image feature vectors
- NLG
  - Templates or trees
  - Midge <noun, verb, preposition>

    + tree substitution grammar + ‘hallucinate’
  - Caption retrieval and recombination
  - Eval: Correctness, order, human likeness, main aspects

(a) The man at bat readies to swing at the pitch while the umpire looks on (Human-authored caption from the MS-COCO dataset Lin et al., 2014)

(b) This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant (Kulkarni et al., 2011)

(c) A person is playing a saxophone (Elliott & De Vries, 2015)

(d) A bus by the road with a clear blue sky (Mitchell et al., 2012)

(e) A bus is driving down the street in front of a building (Mao et al., 2015a)

(f) A gecko is standing on a branch of a tree (Hendricks et al., 2016b)

Building NLG Systems: Knowledge

- Need knowledge
  - Which patterns most important?
  - What order to use?
  - Which words to use?
  - When to merge phrases?
  - How to form plurals
  - Etc
- Where does this come from?

TU Delft

Knowledge "How Was School Today"

Tintarev, Nava, et al. "Personal storytelling: Using Natural Language Generation for children with complex communication needs, in the wild…." *International Journal of Human-Computer Studies* 92 (2016): 1-16.

- How the school works
- How the time tables work
- How mobile the children are
- How much control is useful for them
- … we spent a lot of time at the school!

TUDelft

Scuba: Corpus

- See which patterns humans mention in the corpus, and have the system mention these
- See the words used by humans, and have the system use these as well
- etc

Commercial NLG

- Arria/Datatext: U Abdn spinout company
  - Monitoring equipment on oil platforms
  - weather forecasts
  - Agricultural information
  - Financial summaries

TU Delft

Others

- **Narrative Science** - Builds bespoke “automatic narrative generation” systems
  - Academic roots in computational creativity
- **Automated Insights** - writes “insightful, personalized reports from your data”
  - Non-academic roots
- **Yseop** - “Smart NLG” software that “writes like a human”
  - Chief scientist, Alain Kaeser did NLG in 1980s

TUDelft

Others

- Lots of small young startups, e.g.,
  - **OnlyBoth** “Discovers New Insights from Data. Writes Them Up in Perfect English. All Automated”
  - **InfoSentience** “Developers of the Most Advanced Automated Narrative Generation Software”
  - **Text-on** (German) “Aus abstrakten Daten werden so Texte”
- NLG projects at large companies.
  - E.g., Thomson-Reuters, Agfa
  - More secretive

TUDelft

Common Themes

- Almost all claim to generate narratives/stories from data
- Financial reporting is most commonly mentioned use
- Companies still quite small, but growing field!
  – Fewer than 100 employees, compared to 12,000 at Nuance or 400,000 at IBM
  – But large compared to earlier NLG companies
  – Also lots of them!

Open challenges

- Generating summaries from social media (deal with unstructured language, irony etc)
- Situated language generation (c.f., GIVE challenge, gaming)
- NLG from ontologies/structured knowledge
- Data-driven not so well used in industry

TU Delft

Today

Natural language generation

- NLG pipeline
  - Document planning
  - Microplanning
  - Realisation
- Example systems
  - e.g., ScubaText, HWST

Next time

NLP annotations

TUDelft

Questions?

P3: Thumbs up?: sentiment classification using machine learning techniques

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002.

Review P3: handed out February 28, due **March 6**.

TUDelft