Natural Language Generation and Human Language Production:
a history and an opportunity

David M. Howcroft (@_dmh)
(formerly of) Heriot-Watt University
(soon of) Edinburgh Napier University

Engineering Conversation Workshop, 2021

Overview

What are we talking about? (NLG & HLP)

Where have they connected? (Structures & referring expressions)

What is the current state-of-the-art in NLG? (Rules & neural networks)

Opportunities to connect

Natural Language Generation

Broad definition: computer generated text, whether spoken or written

Most Relevant

- **Data-to-text**: from tables, knowledge graphs, sensor logs, (pseudo) semantic meaning representations, etc, to text
- **Text-to-text**: including summarisation, machine translation, simplification, style transfer

```
name[Aromi], food[Chinese], customer rating[5 out of 5], area[city centre]
```

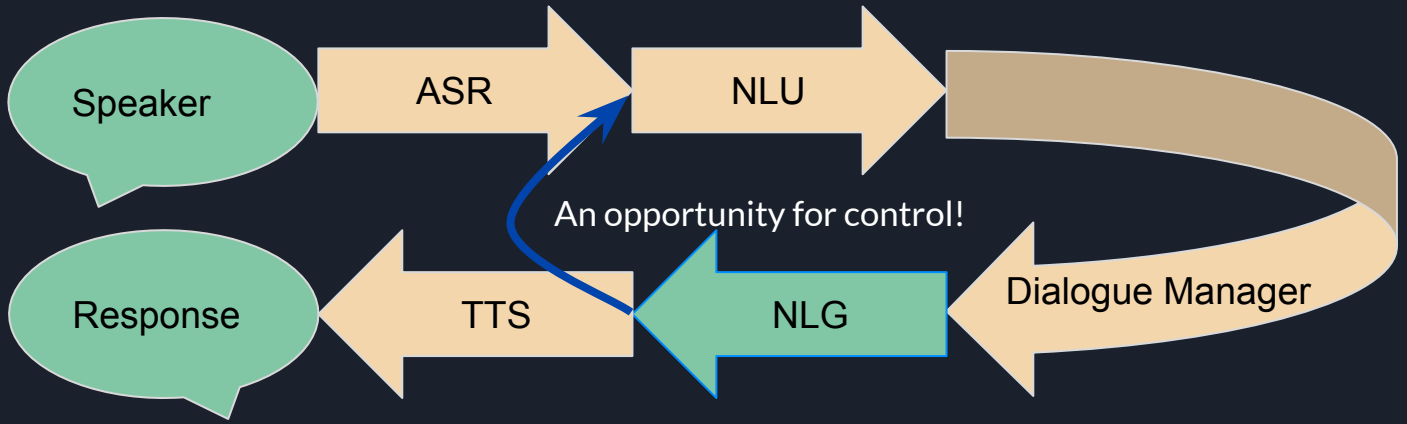
Aromi is a restaurant providing Chinese food. It is located in the city centre. Its customer rating is 5 out of 5.

There's a Chinese place in the city center called Aromi with a 5-star rating.

Five out of five is Aromi, a centrally located Chinese restaurant.

You can't go wrong with Aromi, serving Chinese food in the city centre. 5/5

NLG in (Spoken) Dialogue Systems



Human Language Production

Broad perspective: the realm of psycholinguistics

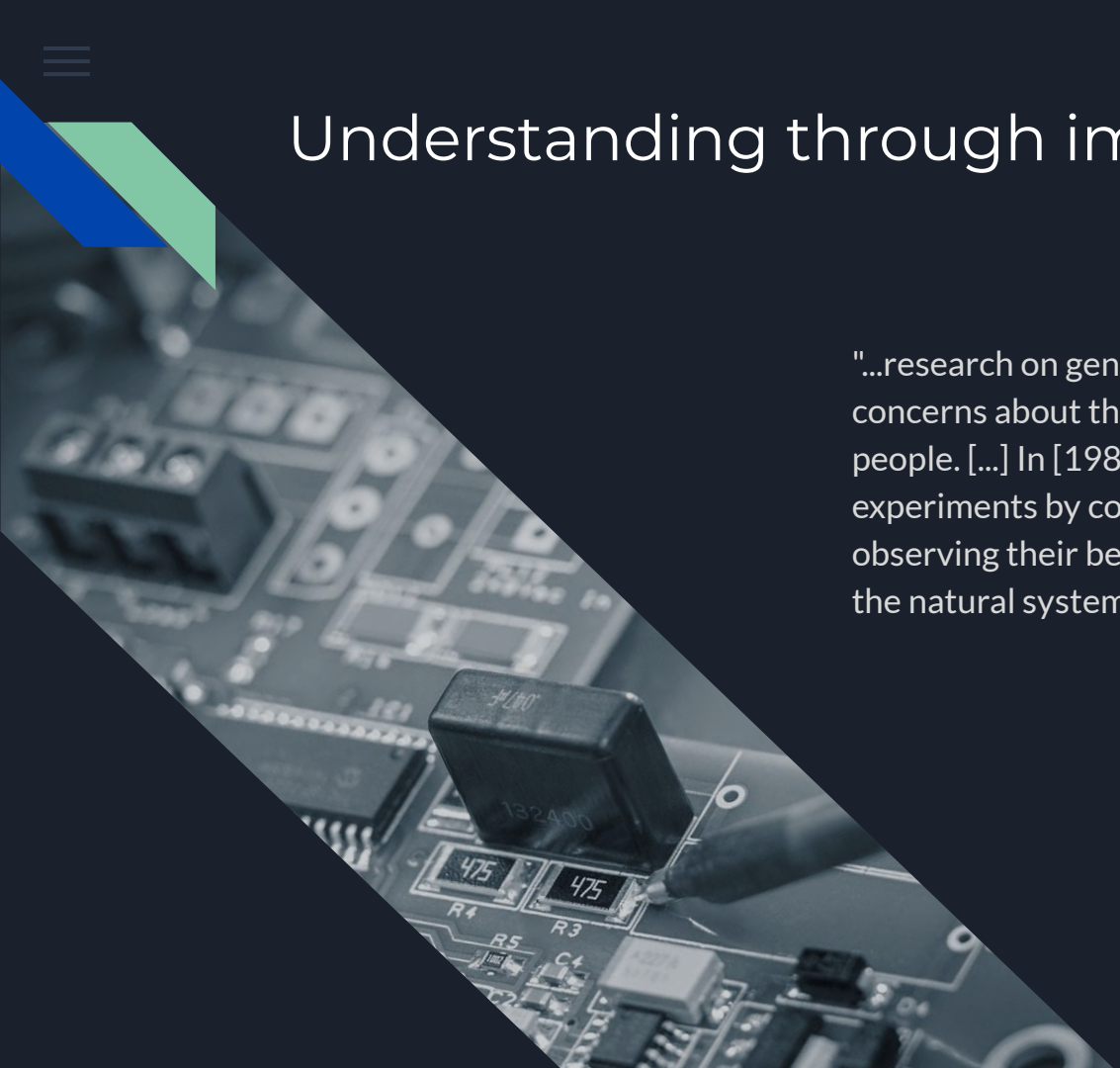
- Cognitive and communicative considerations (psycho)
- Structural and theoretical considerations (ling)

Where do NLG & HLP connect?

01 Understanding structure through implementation

02 Parallel developments with referring expressions

Understanding through implementation



"...research on generation is often aimed at purely scientific concerns about the nature of language and language use in people. [...] In [1980s] AI studies of generation, one experiments by constructing artifacts (computer programs) observing their behavior, and comparing it to the behavior of the natural system under study."

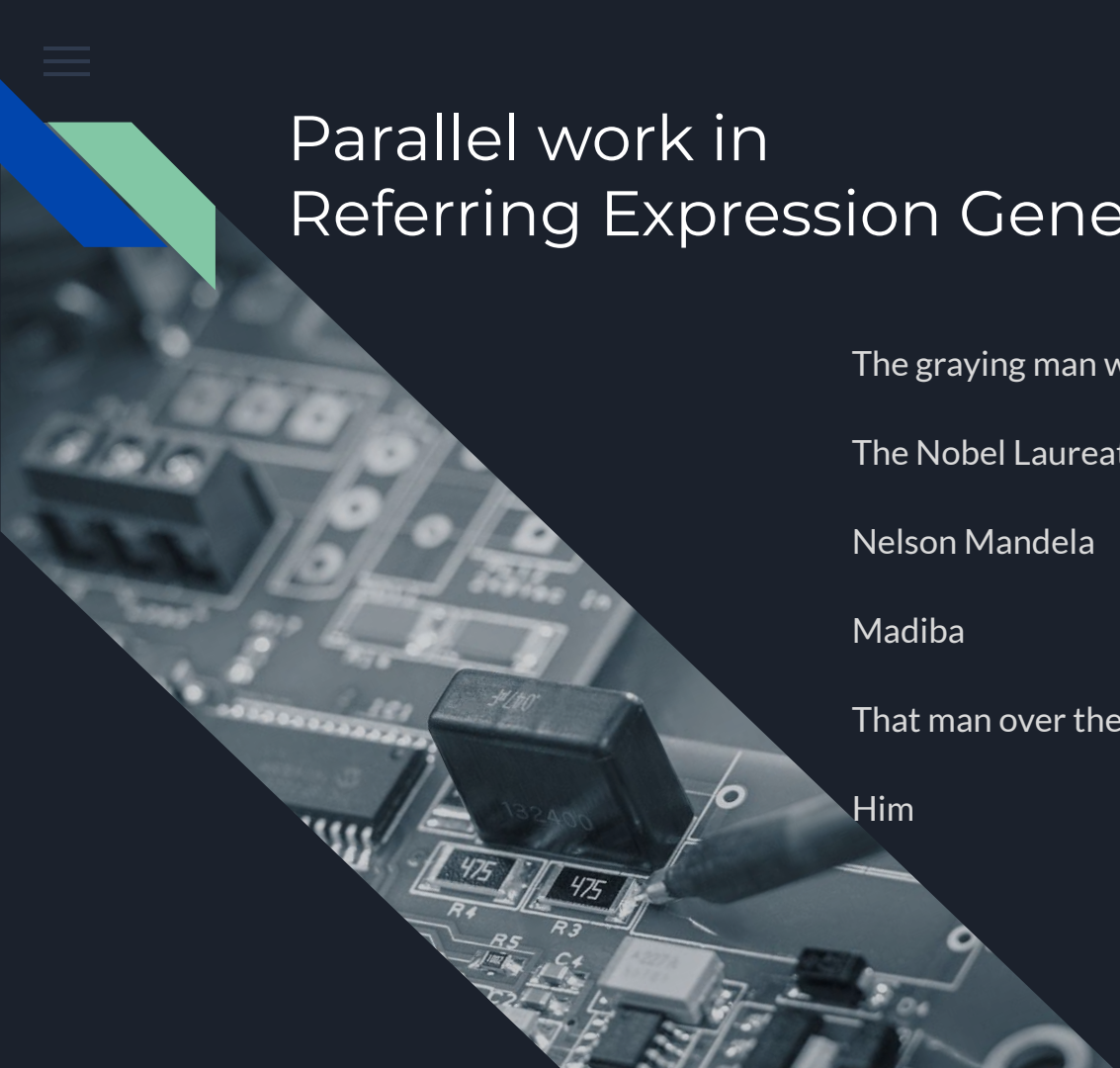
McDonald (1988)

Systemic Functional Linguistics and Natural Language Generation

- SFL focuses on *paradigmatic choices*, providing a natural parallel to NLG
- Early descriptions (e.g. Halliday 1956) relating SFL to machine translation
- Penman (Matthiessen 1981, Mann 1985), SLANG (Patten 1988), & other 'sentence generation' systems developed using SFL in 1980s
- But implementing so many choice points leads to a combinatorial explosion!
- Efforts at formalization (for NLG) contributed to identifying similarities to head-driven phrase structure grammar (HPSG) & other feature-based grammatical formalisms.

"It is now clear that both the paradigmatic orientation of systemic grammar can be usefully complemented by a stronger degree of syntagmatic description and that the structural approaches can benefit from a stronger paradigmatic orientation." –Bateman (1997)

Parallel work in Referring Expression Generation



The graying man with a blue shirt and suit jacket

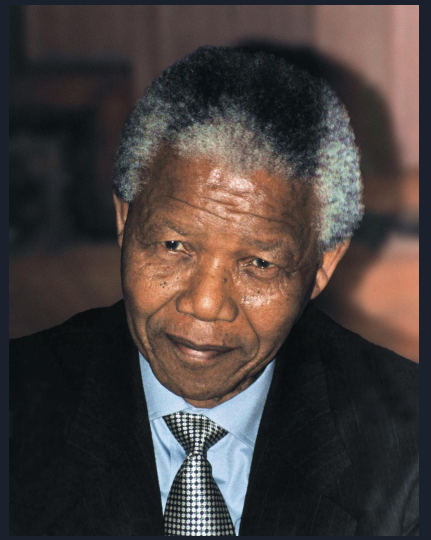
The Nobel Laureate former leader of my birth country

Nelson Mandela

Madiba

That man over there

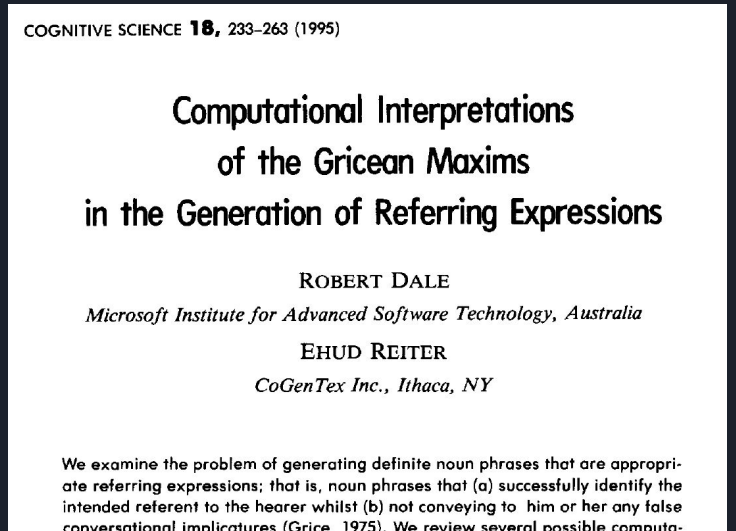
Him



Referring Expression Generation

NLG: greater focus on algorithmic implementation

HLP: greater focus on ego- vs. allo-centric production



COGNITIVE SCIENCE **18,** 233–263 (1995)

Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions

ROBERT DALE
Microsoft Institute for Advanced Software Technology, Australia
EHUD REITER
CoGenTex Inc., Ithaca, NY

We examine the problem of generating definite noun phrases that are appropriate referring expressions; that is, noun phrases that (a) successfully identify the intended referent to the hearer whilst (b) not conveying to him or her any false

ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Memory and Language 54 (2006) 554–573
Journal of Memory and Language
www.elsevier.com/locate/jml

Do speakers and listeners observe the Gricean Maxim of Quantity?

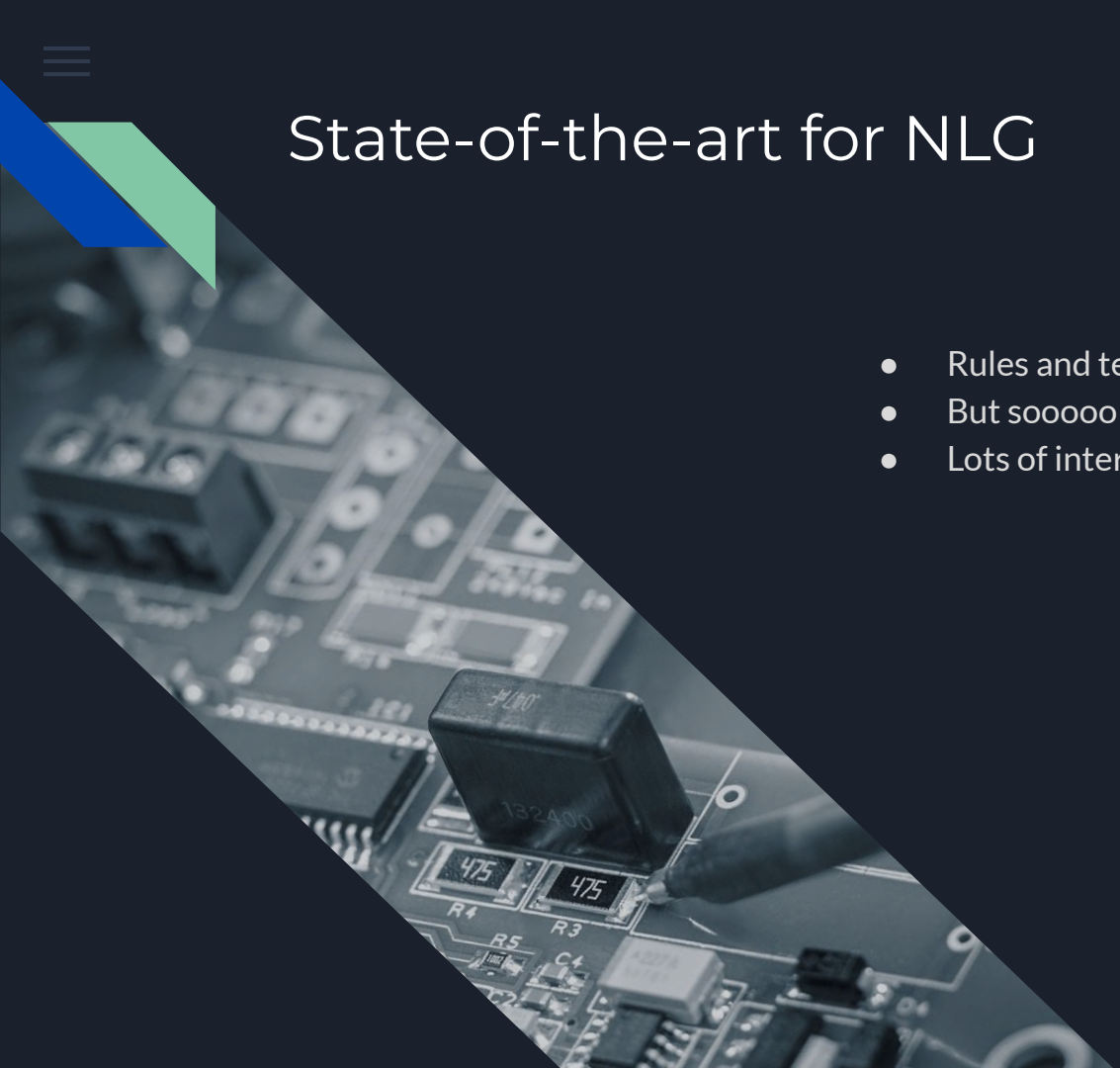
Paul E. Engelhardt [a,*], Karl G.D. Bailey [b], Fernanda Ferreira [a]
[a] *Department of Psychology, Michigan State University, East Lansing, MI 48824-1116, USA*
[b] *Andrews University, Berrien Springs, MI 49104, USA*

Received 25 April 2005; revision received 12 December 2005
Available online 21 February 2006

Abstract

The Gricean Maxim of Quantity is believed to govern linguistic performance. Speakers are assumed to provide as much information as required for referent identification and no more, and listeners are believed to expect unambiguous but concise descriptions. In three experiments we examined the extent to which naïve participants are sensitive to the Maxim of Quantity. The first was a production experiment which demonstrated that speakers over-describe almost one-third of the time. The second experiment showed that listeners do not judge over-descriptions to be any worse than concise expressions. The third experiment used the Visual World Paradigm to assess listeners' moment-by-moment

State-of-the-art for NLG



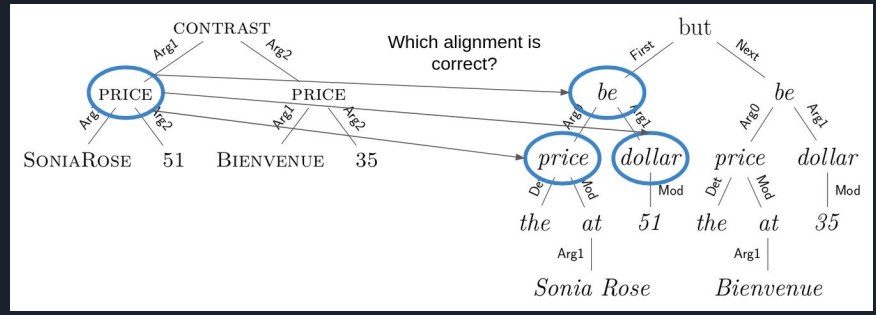
- Rules and templates are still useful
- But sooooo many neural networks in research!
- Lots of interest in improving evaluation

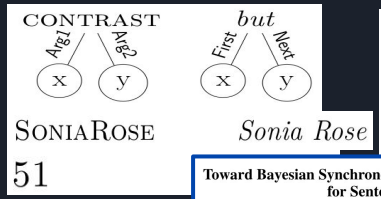
Rules and Templates for NLG

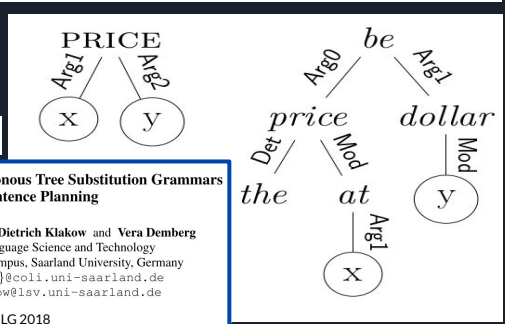
Neural networks lack control.

https://ehudreiter.com/2020/12/01/dont-fixate-on-end-to-end-neural/

But there's more to machine learning than neural networks







Neural network approaches

Not like earlier work building on connectionism (e.g. Kukich 1987)

Increased compute resources → the rise of neural networks for machine learning (2010s)

Two major themes:

- Encoder-decoder models (esp. seq2seq)
- Fine-tuning and/or building on large neural language models

Sequence-to-sequence (seq2seq)

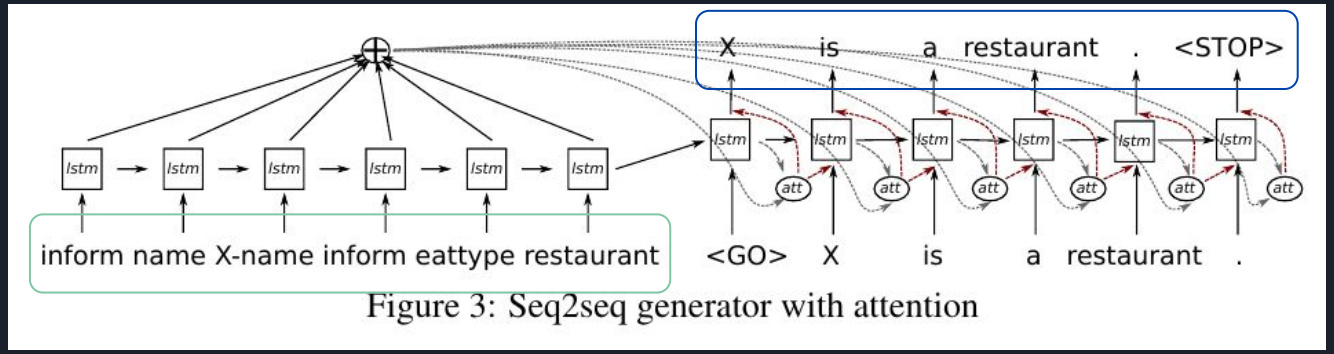


Figure 3: Seq2seq generator with attention

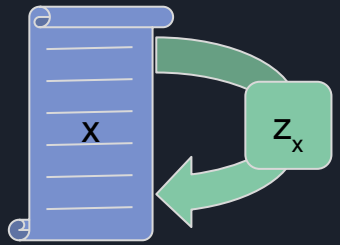
Sequence-to-Sequence Generation for Spoken Dialogue via Deep Syntax Trees and Strings

Ondřej Dušek and **Filip Jurčíček**
Charles University in Prague, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Malostranské náměstí 25, CZ-11800 Prague, Czech Republic
{odusek,jurcicek}@ufal.mff.cuni.cz

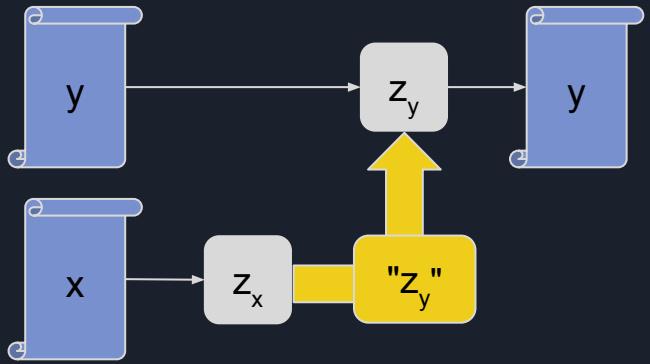
Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 45–51,
Berlin, Germany, August 7-12, 2016.

Adapting Large Pre-trained Language Models

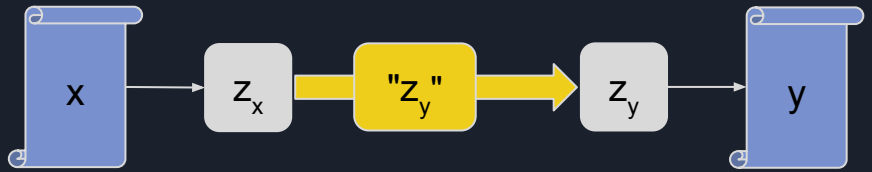
Pretrain on *lots* of text



Train a model to map z_x to z_y



Apply the model to new inputs! (Generate)



Plug and Play Autoencoders for Conditional Text Generation

Florian Mai[†♠] Nikolaos Pappas[♣] Ivan Montero[♣]
Noah A. Smith[♣◇] James Henderson[†]

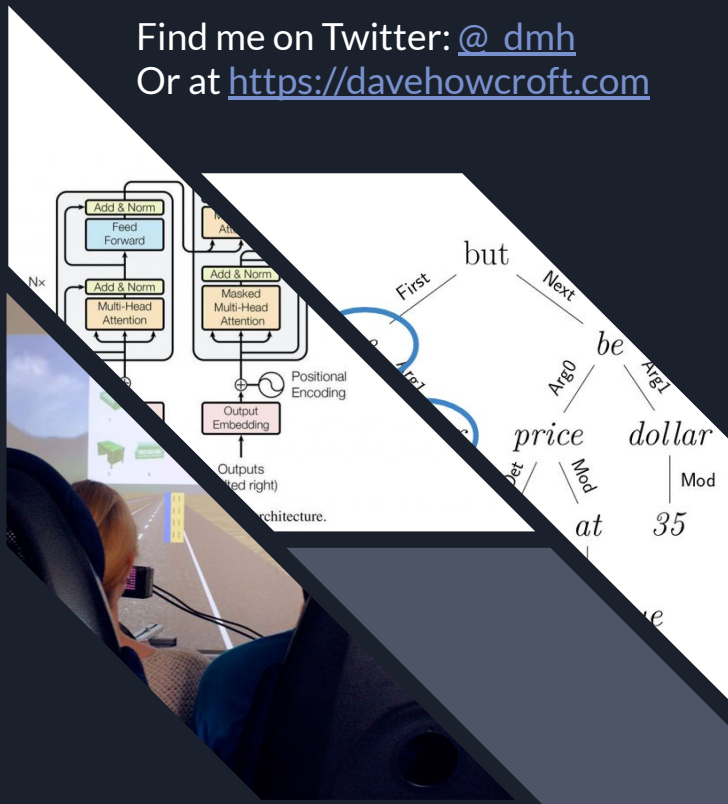
[†]Idiap Research Institute, Martigny, Switzerland
[♠]EPFL, Lausanne, Switzerland
[♣]University of Washington, Seattle, WA, USA
[◇]Allen Institute for Artificial Intelligence, Seattle, WA, USA
{florian.mai,james.henderson@idiap.ch}
{npappas,ivamon,nasmith@cs.washington.edu}

Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 6076–6092,
November 16–20, 2020.

Opportunities

- Joint workshops on REG?
- Using large LMs as a kind of control theory forward model for NLG
- Bringing more structures and features back into NLG (cf. Balakrishnan et al. 2019 or work from Marilyn Walker's group)
- Collaborations with chatbot / SDS researchers on modelling context
  - e.g. applying the rational speech act model (cf. Shen et al. 2019)

Find me on Twitter: @_dmh
Or at https://davehowcroft.com



Thank you for your attention! Let's discuss :D

Where do NLG & HLP connect?

01 Understanding through implementation (SFL & NLG)

02 Parallel developments (REG)

03 Comparisons to humans (NLG Evaluations)