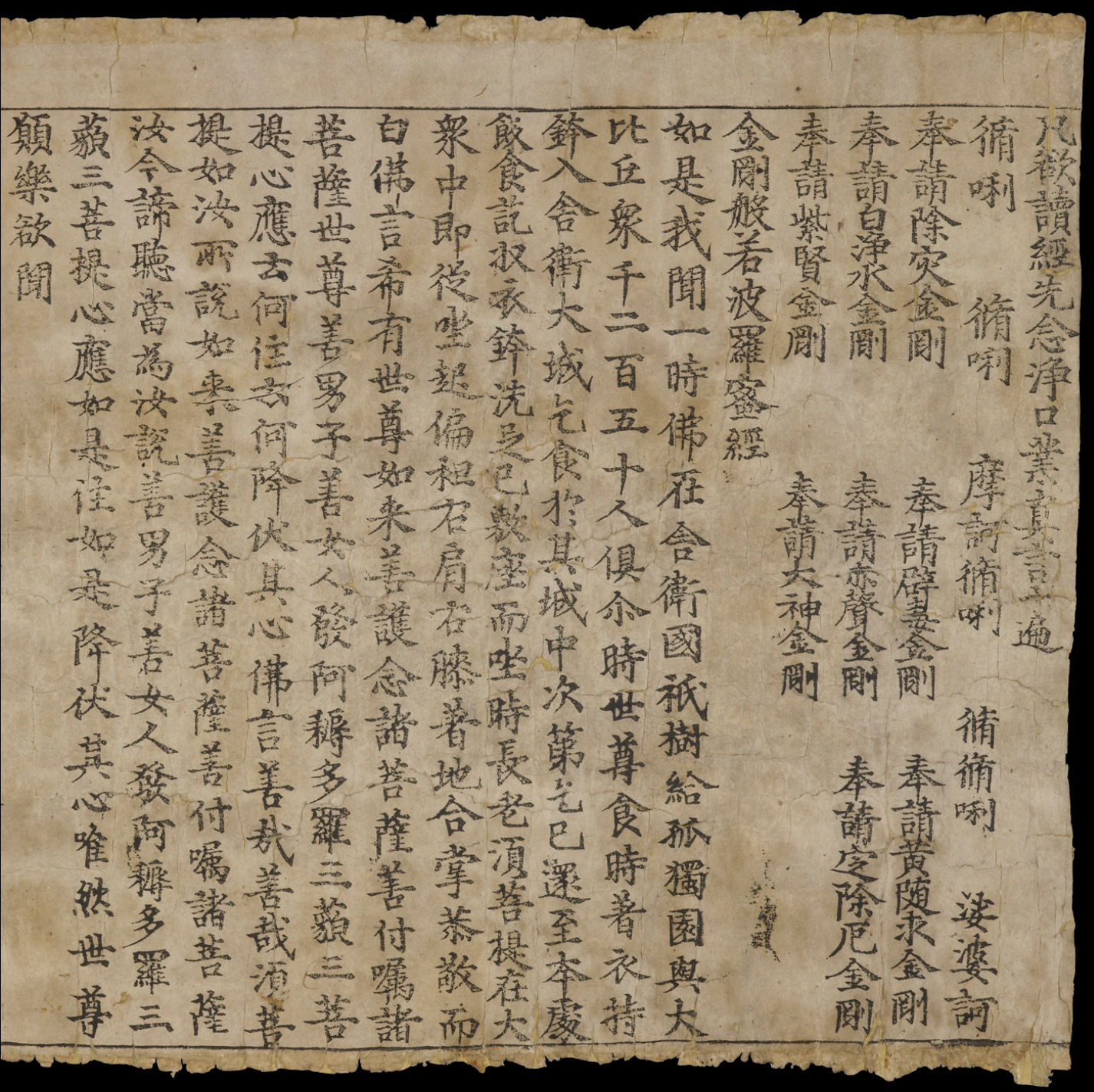
Neural Machine Translation

Recent Advances and Remaining Challenges

Hagen Fürstenau

Amazon Development Center Germany (Berlin)


凡欲讀經先念淨口業真言七遍
脩唎 脩唎 摩訶脩唎 娑婆訶
奉請除災金剛 奉請辟毒金剛 奉請黃隨求金剛
奉請白淨水金剛 奉請赤聲金剛 奉請定除厄金剛
奉請紫賢金剛 奉請大神金剛
金剛般若波羅蜜經
如是我聞一時佛在舍衛國祇樹給孤獨園與大
比丘衆千二百五十人俱尒時世尊食時著衣持
鉢入舍衛大城乞食於其城中次第乞已還至本處
飯食訖收衣鉢洗足已敷座而坐時長老須菩提在大
衆中即從坐起偏袒右肩右膝著地合掌恭敬而
白佛言希有世尊如來善護念諸菩薩善付囑諸
菩薩世尊善男子善女人發阿耨多羅三藐三菩
提心應云何住云何降伏其心佛言善哉善哉須菩
提如汝所說如來善護念諸菩薩善付囑諸菩薩
汝今諦聽當為汝說善男子善女人發阿耨多羅三
藐三菩提心應如是住如是降伏其心唯然世尊
願樂欲聞

Earliest completely surviving and dated printed book (868 CE)

Buddhist Diamond Sutra

Translation from Sanskrit into Classical Chinese by Kumārajīva (344–413 CE)

[T]ranslation [...] is in fact an art both estimable and very difficult, and therefore is not the labor and portion of common minds

– Ignacy Krasicki (1735–1801 CE)

Translation

- From a *source* language into a *target* language

- Literal translation vs. free translation
  - Fidelity: Accurate rendering meaning of the source text
  - Transparency: Syntactic and idiomatic naturalness in the target language

- Fidelity and transparency often at odds

[W]hat was beautiful in the Greek or Latin, would not appear so shining in the English
– John Dryden (1631–1700 CE)

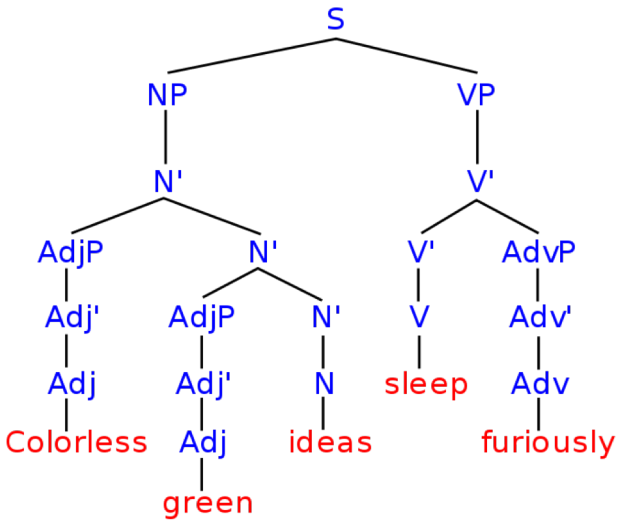
Computers and Decoding

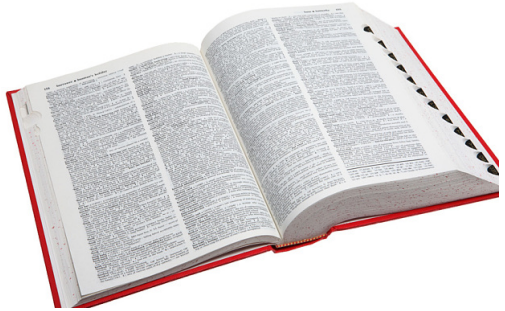
- World War II saw the rise of the digital computer
  - Successful decryption of German military communication ("Enigma" machine, Bletchley Park)

- Memorandum "Translation" by Warren Weaver (1949)
  - Ideas for applying computers to the problem of translation
  - How to go beyond word-by-word translation? → Transparency
  - Identified problem of *ambiguity* and context → Fidelity
  - Suggested applying techniques from cryptography

Rule-based Machine Translation

1. Analyze grammar of source sentence
2. Map words/concepts to target language
3. Generate sentence according to grammar of target language





Farblose grüne Ideen schlafen wütend.

Rule-based Machine Translation

Advantages

- Straightforward application of computers to the task
- Easy to explain outputs
- Full manual control over lexicon and rules

Disadvantages

- Requires manually curated resources (lexicon, translation rules)
- Manual effort for extending/adapting to new domains
- High-quality rule systems can become very complex

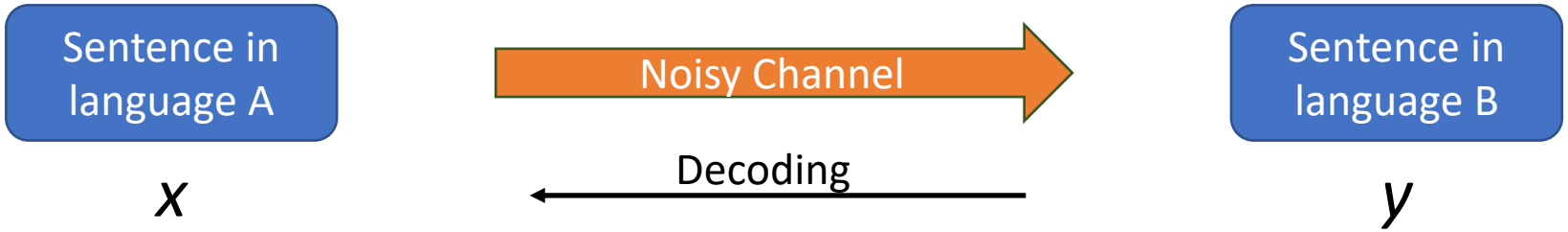
Statistical Machine Translation

Cryptography and Machine Translation:

> *[O]ne naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."*
>
> *– Warren Weaver, 1949*

Statistical Noisy Channel Model:

Statistical Machine Translation



Recover *x* from observed *y*:

$$P(x \mid y) \propto P(x)\,P(y \mid x) \quad \text{(Bayes' Theorem)}$$

Language Model: Characterize likely target language sentences

- Transparency / **Fluency**

Translation Model: Model translation process

- Fidelity / **Adequacy**

Statistical Machine Translation

- Statistical models estimated from *data* (large text corpora)

- Language Model
  - Probability (“fluency”) of a target language sentence
  - Estimated from large target language corpus
  - N-gram models: Based on word frequencies (words, word pairs, …)

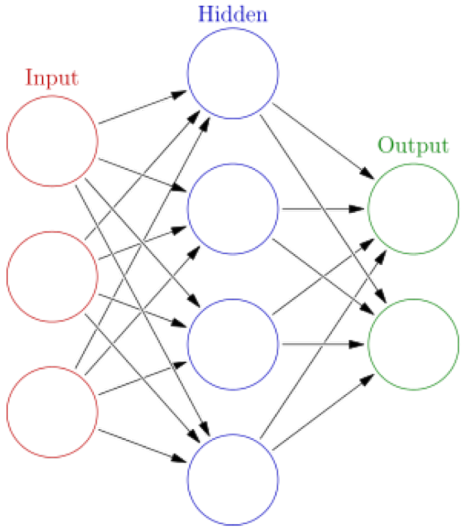
- Translation Model
  - Probability (”adequacy”) of a translation given the source sentence
  - Estimated from bi-texts (e.g. multilingual parliament proceedings)
  - Statistical phrase dictionaries: Translations and probabilities

Statistical Machine Translation

- State of the art in machine translation until a few years ago
- Outputs often intelligible but still awkward
  - Wrong grammar
  - Non-idiomatic expressions (too literal)
  - Untranslated source words
- Translation quality seemed to be reaching a plateau

Neural Networks

- Inspired by neurons in the brain
  - Each neuron takes inputs (numbers)
  - Computes a non-linear function
  - Sends output to other neurons



- Example: Computer vision / Object recognition
  - Input neurons: Pixels of an image
  - Output neurons: One of a set of object classes (e.g. cat, dog, truck, …)
  - Superior to previous approaches

Neural Networks in Statistical MT

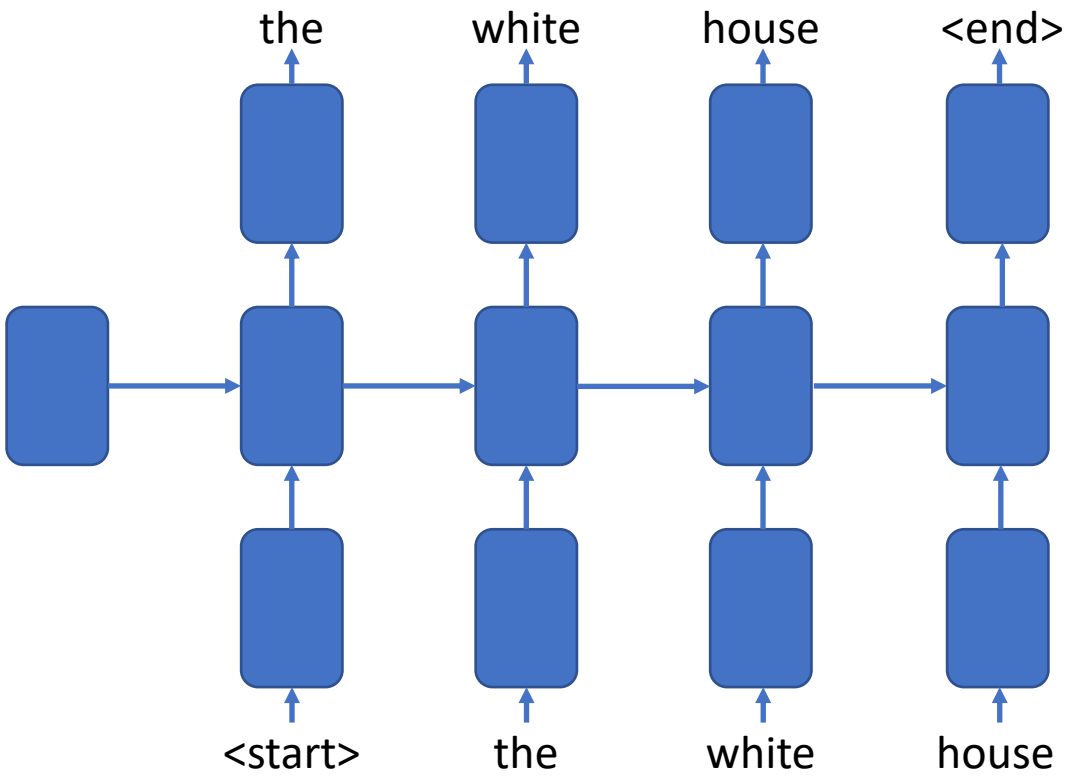
- First approaches: Put neural networks into Statistical MT system

- Translation Model
  - Probability of target sentence, given source sentence
  - Computed from lots of features
    - How often was a phrase translated this way in training corpus?
    - How often were the individual words translated this way?
    - What context words are there?
    - **What translation words does a neural network model propose?**

- Language Model

Neural Network Language Models

- Language model rates fluency of sentences
  - Idea: Use a recurrent neural network (RNN) to predict the next word



Neural Network Language Models

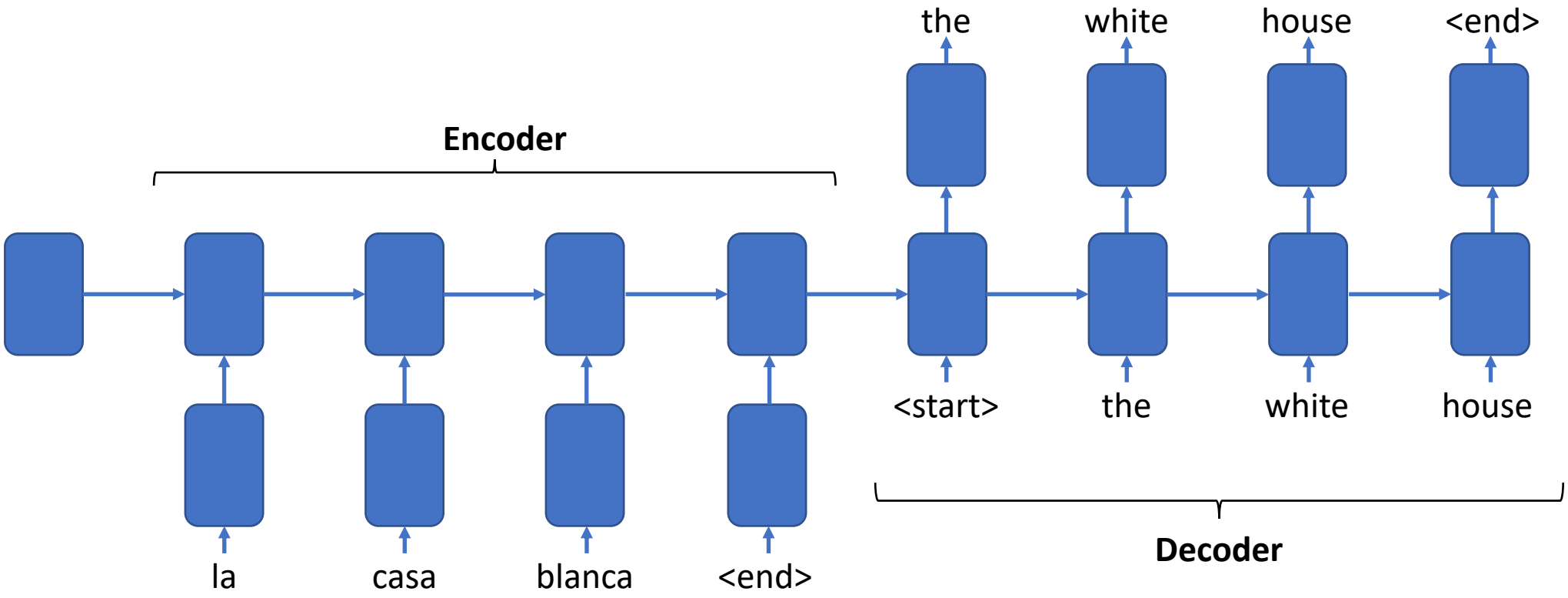
Neural Language Model can be used to:

1. Compute P(x), i.e. score a sentence word-by-word
   - At each position, compute $P(x_i \mid x_{<i})$, multiply together for P(x)
   - In contrast to n-gram language models, uses *unbounded history*!

2. Generate a likely sentence
   - At each position, sample next word from $P(x_i \mid x_{<i})$
   - Not so useful for simple language models, but…

Encoder-Decoder Neural MT Architecture

With a small change, this can be turned into a translation system:



Decoding the Translation

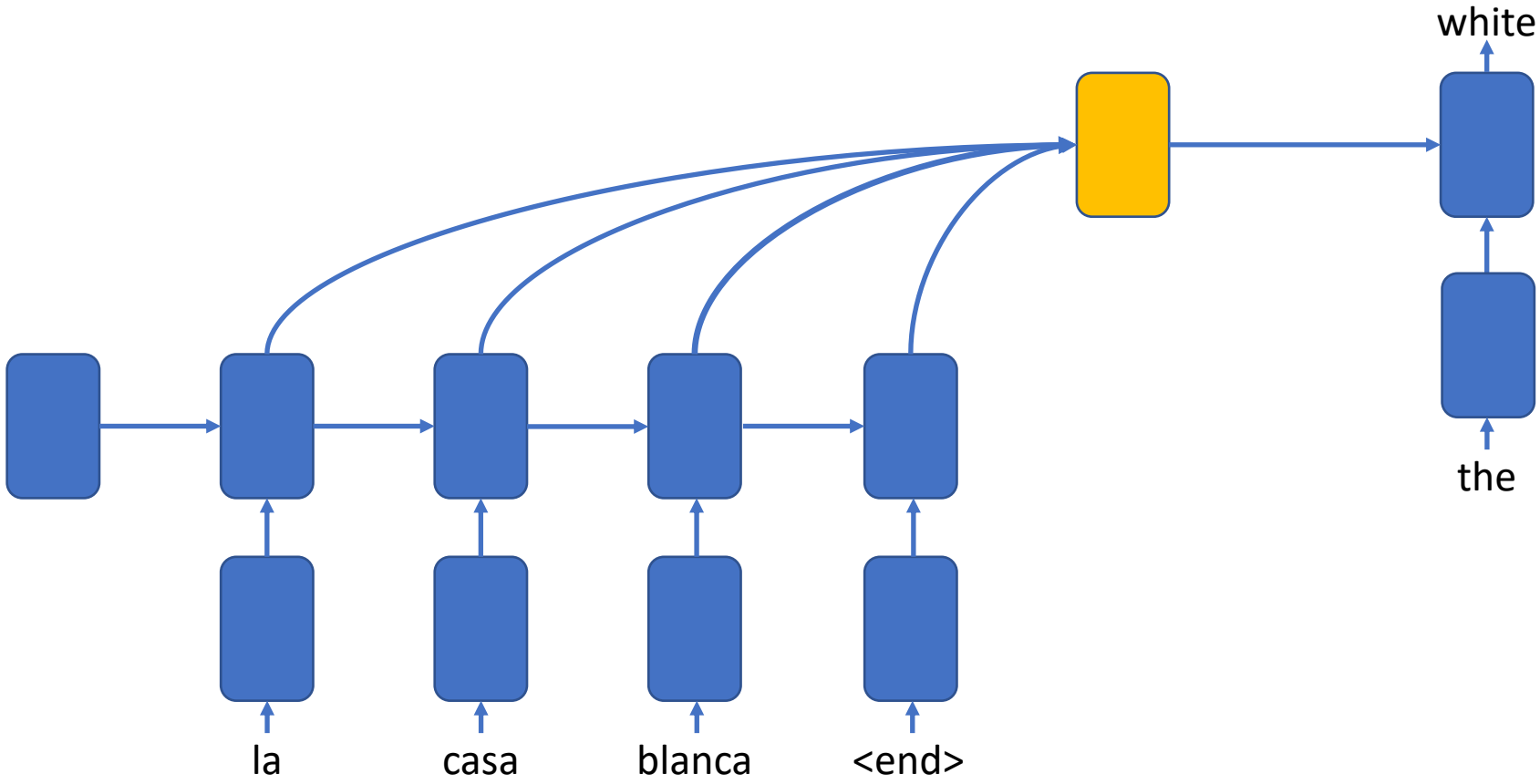
- Encoder-decoder model is a conditional language model
  - Conditioned on the input received from the encoder
  - Helps fluency of the output

- Decoding: Generate translation word by word
  - Greedy search: Pick most probable next word, i.e. maximize $P(x_i \mid x_{<i})$
  - Beam search: Maintain “beam” of the k best hypotheses so far

- Problem: Information bottleneck between encoder and decoder
  - Quality deteriorates for longer sentences

Neural MT with Attention

Paying more or less attention to parts of the source sentence:



Attention Mechanism

- At each decoder step, compute a “context vector”
  - Convex combination of the hidden encoder states
  - Weights computed by a neural network

- Use context vector as input to the decoder
  - In addition to the previous word
  - Allows the decoder to “focus on” or “pay attention” to parts of the source

amazon

Neural Machine Translation (RNNs)

- What we saw so far: Recurrent neural networks (RNNs)
  - Usually have multiple layers of neurons in encoder and decoder (“deep networks”)

- RNNs with attention (Bahdanau et al., 2014)
  - Much better translations quality than Statistical MT
  - Especially much more fluent outputs (based on language models)
  - Sometimes show surprising adequacy problems (hallucination)

- Challenges
  - Long time and compute power to train models on data
  - Significantly slower in translation
  - Active research on finding better neural network architectures

amazon

Transformer NMT Model

- Use attention mechanisms everywhere (Vaswani et al., 2017)
  - Multiple encoder and decoder layers
  - Connected by multiple attention computations
  - No recurrent connections (“memory over time”)

- Advantages
  - Faster training
  - Higher translation quality

- Disadvantages
  - Slower translation

amazon

Convolutional Networks for Neural MT

- Use convolutional networks (Gehring et al., 2017)
  - Inspired by computer vision
  - Detect patterns in words

- Faster in training than RNNs
- Competitive translation quality
- Potentially more robust

Summary of Neural MT Models

Most popular approaches at this time:

1. Recurrent Neural Networks with Attention
2. Transformer with (Self-)Attention everywhere
3. Convolutional Neural Networks

Future directions:

- Mix and match elements of these approaches (Domhan, 2018)
- Completely new architecture?

Machine Translation at Amazon

Title

Für größere Ansicht Maus über das Bild ziehen

Rosenthal 10430-800001-26014 Maria Vase 14 cm, weiß

von Rosenthal

Geben Sie die erste Bewertung für diesen Artikel ab

Unverb. Preisempf.: ~~EUR 49,00~~

Preis: EUR 42,38 **kostenlose Lieferung.**

Sie sparen: EUR 6,62 (14%)

Alle Preisangaben inkl. MwSt.

Nur noch 2 auf Lager

Verkauf und Versand durch Neuerts - der Shop für äußerst faire Konditionen für Sammler und Liebhaber von Küchen-/Wohnambiente. Für weitere Informationen, Impressum, AGB und Widerrufsrecht klicken Sie bitte auf den Verkäufernamen.

3 neu ab EUR 40,00

- Qualitätsporzellan der Marke Rosenthal
- Höhe 14 cm
- spülmaschinenfest

Bullet points

Produktbeschreibungen

Der Siegeszug der Form "Maria" begann im Jahr 1916. Benannt wurde das Service nach der Frau des Firmengründers Philipp Rosenthal, der die 35 Jahre jüngere Maria 1916 ehelichte. "Maria" steht für Porzellangeschichte und Tischkultur. Es ist eines der beliebtesten und erfolgreichsten Rosenthal Porzellanservice. Seit Generationen ist es das Porzellan zur Hochzeit und anderen festlichen Anlässen: stilvoll und zeitlos. Diese polygonale Geschirrform mit dem charakteristischen Kantenrelief in Form einer Früchtegirlande wurde zum umfangreichsten Rosenthal-Service. Es sind mehr als 80 verschiedene Einzelteile erhältlich. Über Jahrzente hinweg erfreut sich "Maria" großer Beliebtheit und gehört zu den umfangreichsten und meistverkauften Geschirrformen aller Zeiten. Mit seiner Vielzahl an Spezial- und Zusatzteilen erfüllt "Maria" höchste individuelle Ansprüche an die klassisch zeitgemäße Tischkultur und präsentiert sich als komplettes Gourmet-Programm. Der elegante Charme dieser Form liegt vor allem in der klassizistisch anmutenden Form und der Balance klassischer und romantischer Stilelemente – wie dem zarten Granatapfel-Relief. Die klassizistische, vieleckige Grundform wird von einem zierlichen, feingeschwungenen Granatapfel-Relief eingefaßt, das Gestaltungselemente des Biedermeier aufgreift.

Description

Translating Product Pages

- Amazon serves customers in many countries/languages
- Translate product pages (hundreds of millions already)
  - Offer larger selection to customers
  - Offer more opportunities to sellers
  - Show the website in the customer's language of preference
- Ongoing research on improving translations of product pages

Amazon Translate

Launched translation service on Amazon Web Services (AWS) this year

- Open to all AWS customers
- Supports translation between English and 12 other languages
- Extension to more languages planned

- Based on the latest neural machine translation technology
- Ongoing research to provide the best translation quality

http://aws.amazon.com/translate

Sockeye – Our Neural MT Toolkit

- Implementation of all three major NMT architectures
- Released as open-source last year
  - Ongoing contributions, both from Amazon and the research community
- WMT 2018 Competition
  - Winning German-English system (RWTH Aachen) based on Sockeye

https://github.com/awslabs/sockeye

amazon

amazon research awards

The Amazon Research Awards (ARA) program offers awards of up to $80,000 in cash and up to $20,000 in AWS Cloud Credits to faculty members at academic institutions worldwide for research in the following areas:

- Computer vision
- Economics
- Knowledge management and data quality
- Machine learning algorithms and theory
- Natural language processing
- Operations research and optimization
- Personalization
- Robotics
- Search and information retrieval
- Security, privacy and abuse prevention
- Speech

http://ara.amazon-ml.com

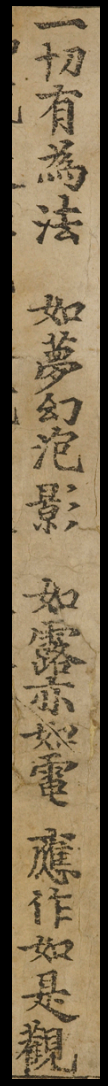
Remaining Challenges

- Neural Machine Translation has been a revolution

- Explainability
  - Systems may start producing nonsense (hallucination, stuttering, …)
  - More research needed on why this happens & how to prevent it

- Robustness
  - Like earlier approaches, still easily confused by spelling mistakes, colloquial language etc.

- Ambiguity & context
  - Many translations come out beautifully, but hard cases remain
  - We translate sentence-by-sentence, but sometimes broader context is needed
  - Some ambiguity is very hard to resolve, even for humans
  - Perfect translation requires deep understanding of the meaning of texts



Diamond Sutra

All conditioned phenomena
Are like a dream, an illusion, a bubble, a shadow
Like the dew, or like lightning
You should discern them like this

– English by A. Charles Muller (from Kumārajīva's Classical Chinese translation)

As stars, a fault of vision, as a lamp,
A mock show, dew drops, or a bubble,
A dream, a lightning flash, or cloud,
So should one view what is conditioned.

– English by E. Conze (from the Sanskrit original)