



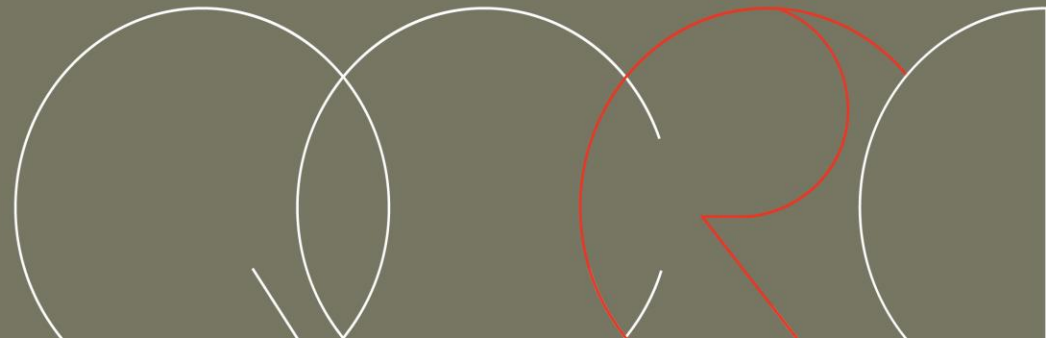
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute



Machine Translation – Winds of Change

Stephan Vogel
Research Director
Arabic Language Technologies

InterACT 25
August 14, 2016
Baden-Baden, Germany



Wind of Change – This one for Alex



The future's in the air
I can feel it everywhere
Blowing with the wind of change

Wind of Change – A rock ballade by the Scorpions 1990
Celebrating the political changes of the times



معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
عضو في مؤسسة قطر
Member of Qatar Foundation



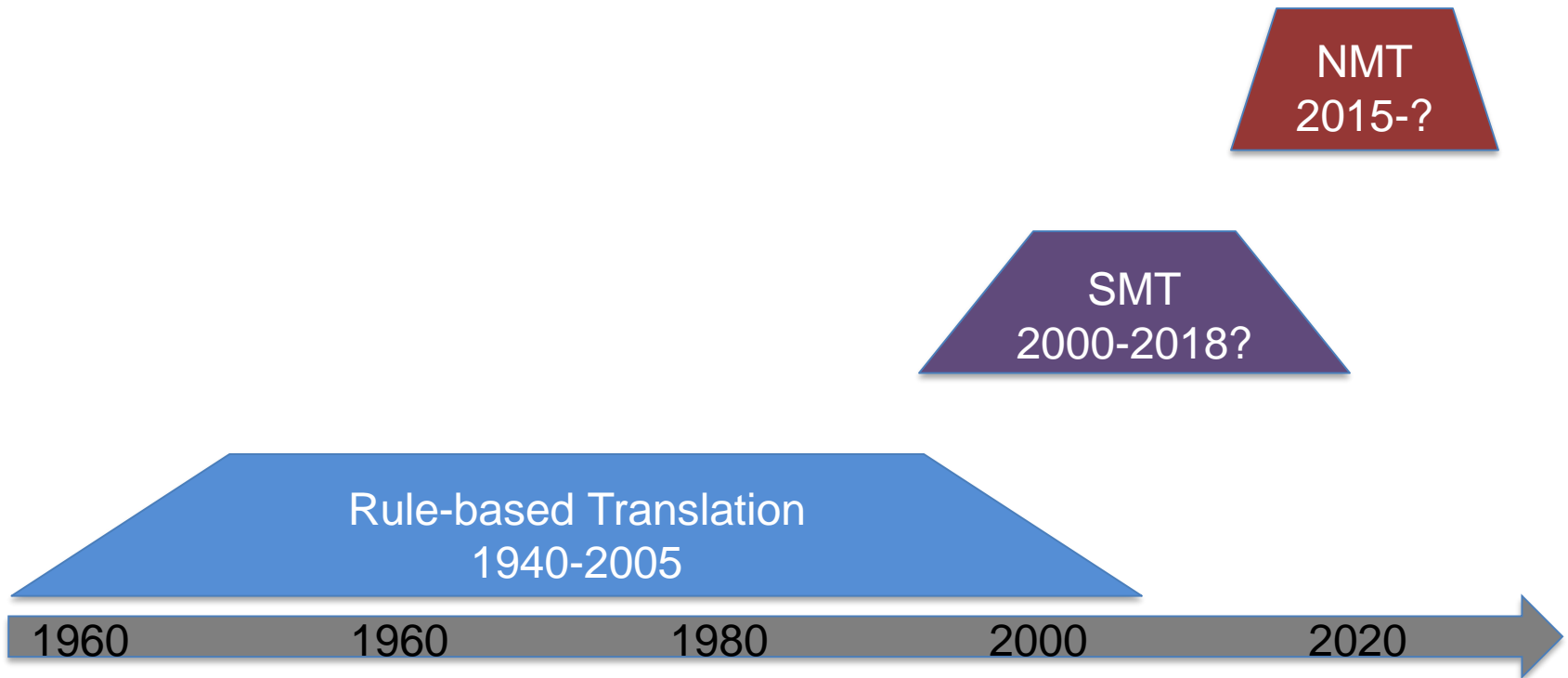
جامعة حمد بن خليفة
HAMD BIN KHALIFA UNIVERSITY

Wind of Change = Paradigm Shift

- Paradigm Shift:
Term introduced by Thomas Kuhn in his famous book “The Structure of Scientific Revolutions” (1962) to describe progress in science
- Paradigm: what the members of a certain scientific community have in common, that is to say, the whole of techniques, practices, patents, and values shared by the members of the community.
- We could also say: Which papers are accepted at conferences
 - Look at recent NAACL and upcoming ACL and you see a paradigm shift happening
- Example: Ptolemaic to Copernican model of the universe



Paradigm Shifts in Machine Translation



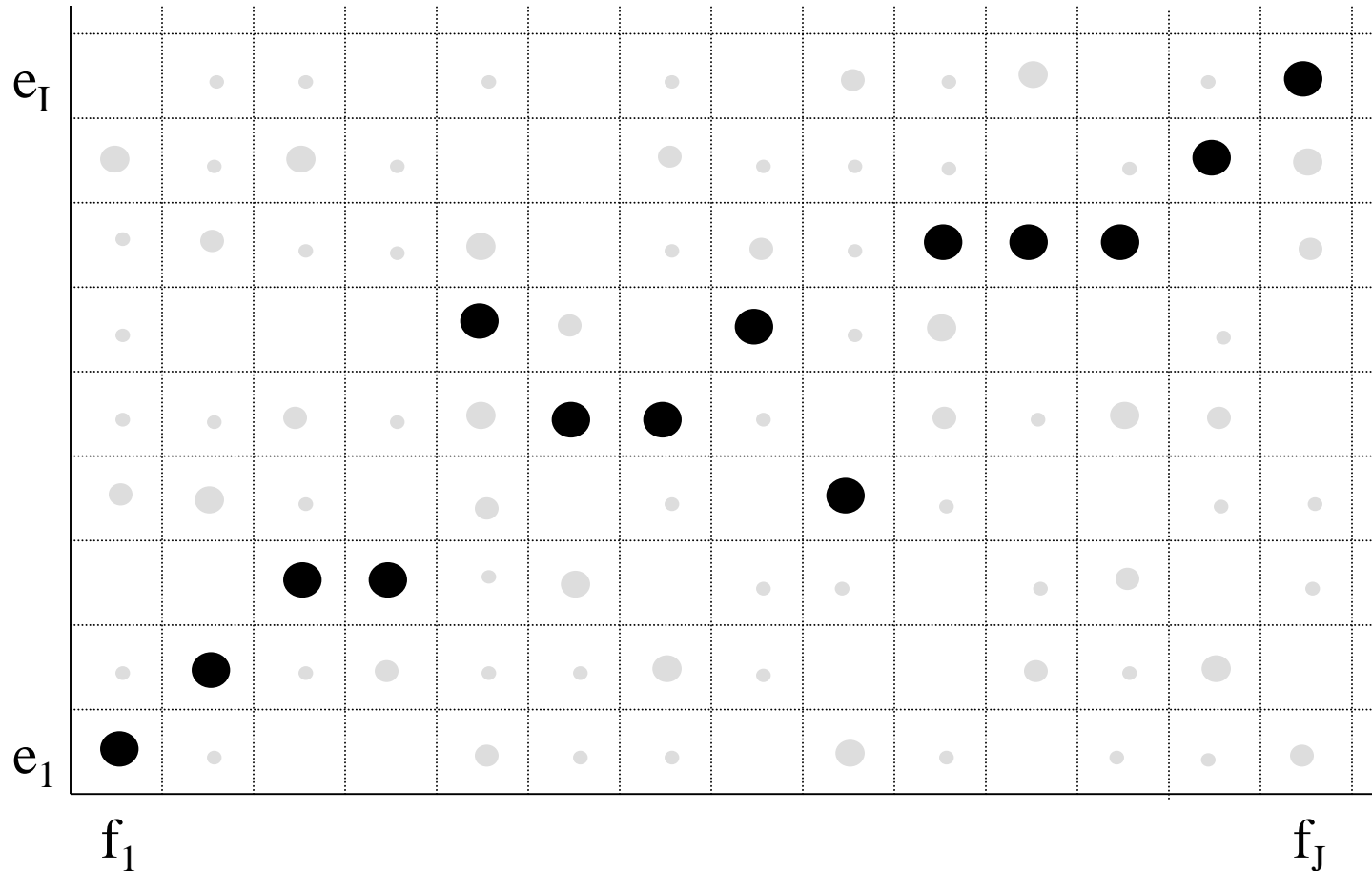
The Old Paradigm: Statistical MT

- Train word alignment
- Extract phrase pairs (or sub-tree to sub-tree pairs)
- Learn word or phrase reordering patterns
- Train n-gram language model for target language
- Use inventory of these translation pairs, reordering patterns, and language model to translate new sentences



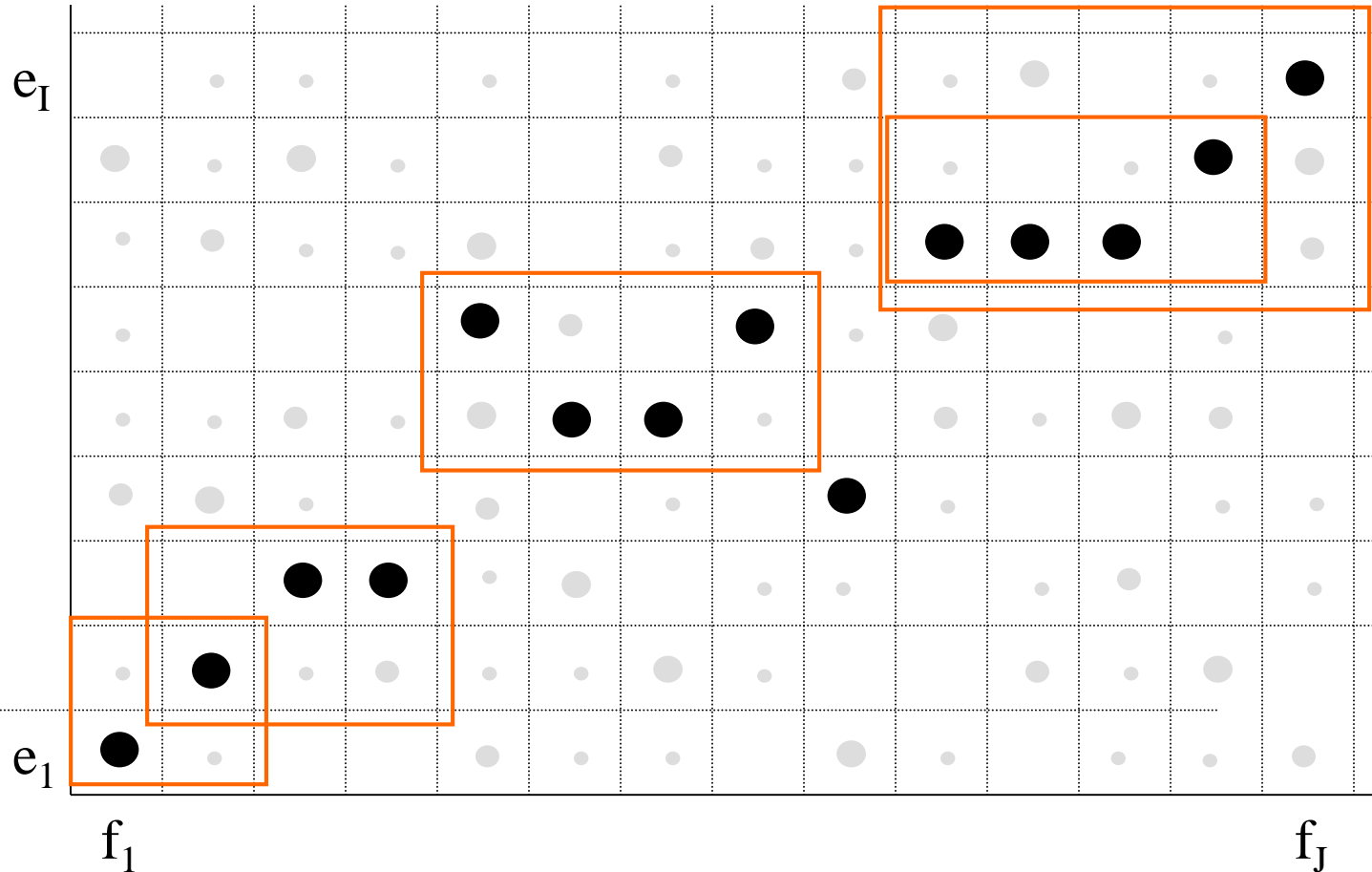
Word Alignment

- Calculate Viterbi path (i.e. path with highest probability)

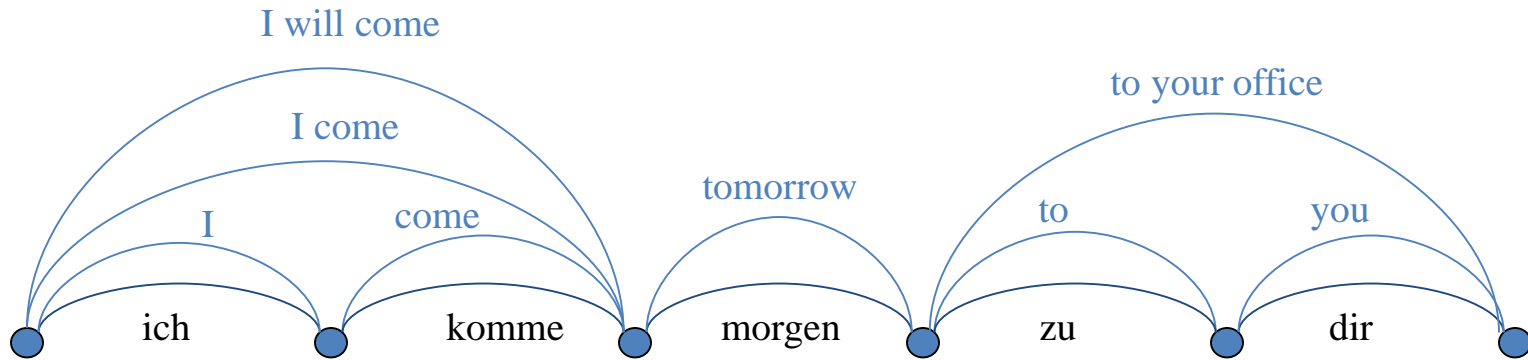


Extract Phrase Pairs

- Read off source phrase – target phrase pairs



Decode New Sentences



h: c=11000, t=I will come

h: c=11011, t=I will come to your office

h: c=11111, t=I will come to your office tomorrow

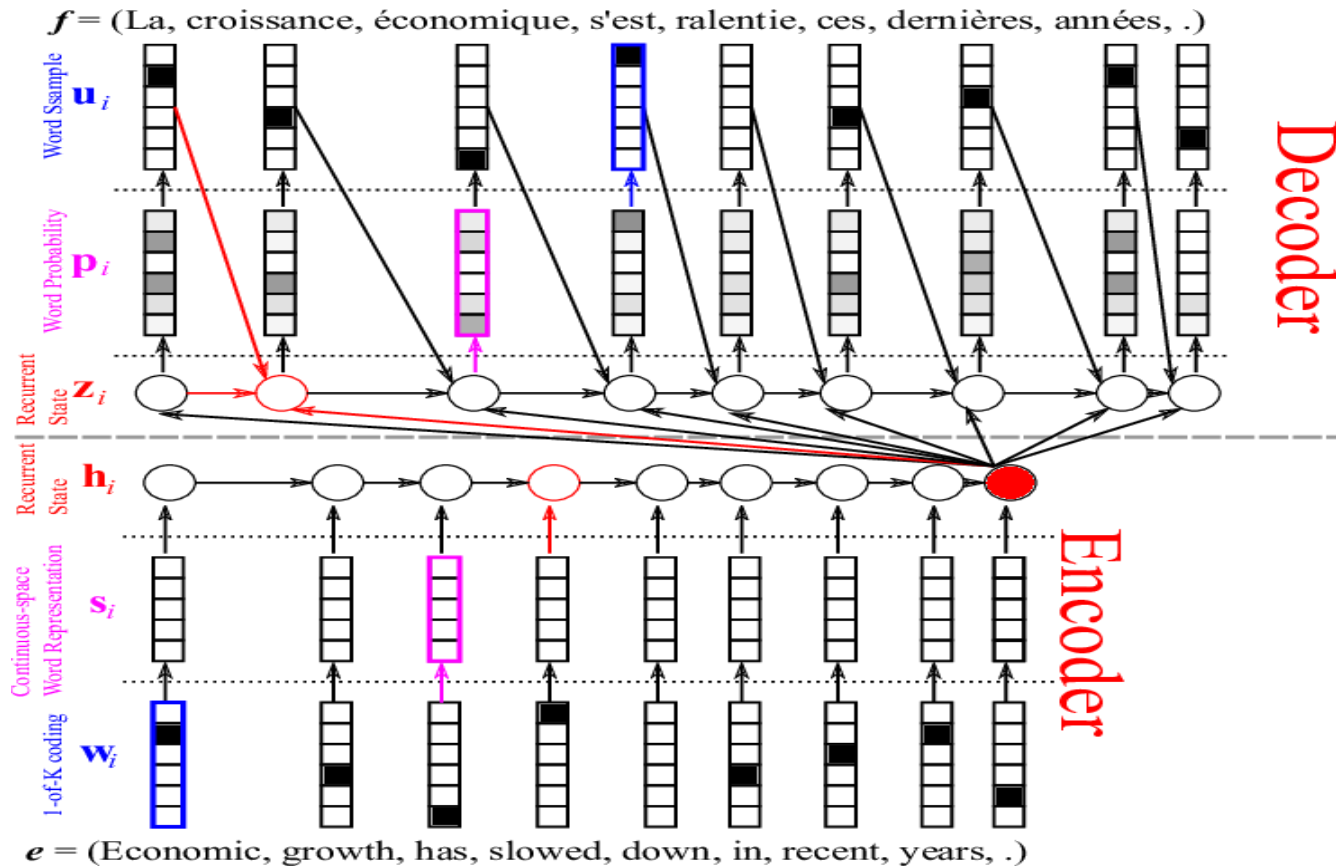


The Old Paradigm: Statistical MT

- Data-Driven:
 - More data -> better translations
 - Low resource scenarios – oh well!
- Serious Problem: Each word is an island
 - Related words (e.g. red, green, blue, ...) are not related in the MT system
 - Worse: different word forms (e.g. table, tables) are complete strangers in SMT
- Some repair mechanisms
 - Factored models
 - Word class models
 - AMR Abstract meaning representation



The New Paradigm: Neural MT



From Kyunghyun Cho: <https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-2/>



The New Paradigm: Neural MT

- Map words into vector space: Embeddings
 - Related words hopefully close in that space
- Process sequence of words and build representation of entire sentence (Encoder)
- Generate sequence of words starting from this representation, also using previously generated words (Decoder)
- NMT is data-driven as SMT
 - Currently no indication that it will solve the low-resource MT problem



Back to the Future: A Few Observations



- Back to the 1990s – word for word translation
 - Right now NMT operates on word level, or sub-word level
 - Can we capture the benefits of phrase-based translation?
 - Can we incorporate syntactic structure?



- No explicit modeling of word alignment
 - Attention mechanism introduced to recover some of the power of the word alignment models
- No specific modeling of word reordering
 - Such models helped a lot in SMT

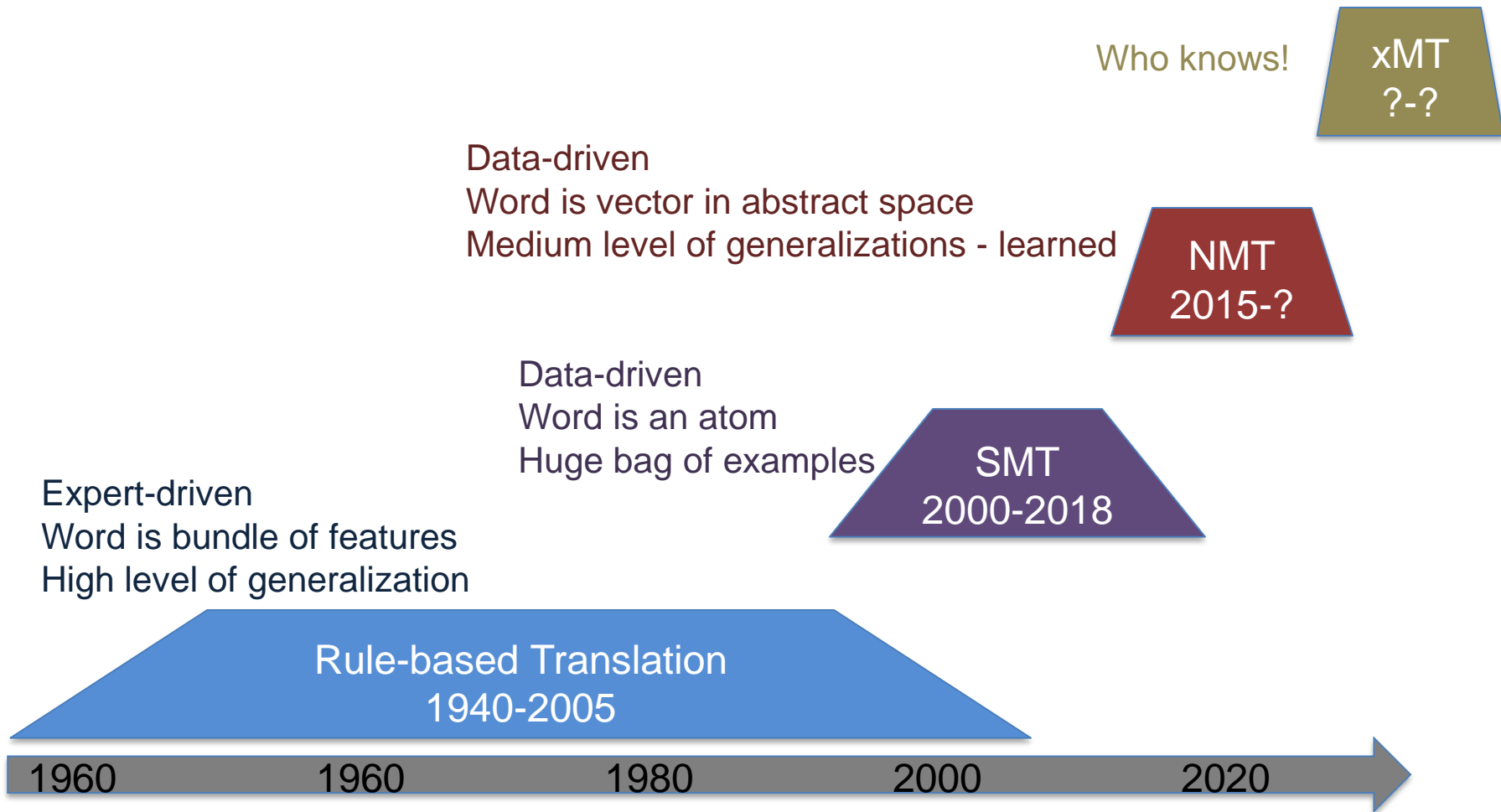


Déjà Vu – Neural Language Processing

- Neural Network based machine learning is not new
 - Created a lot of hype 30 years ago!
 - Rumelhart, D.E., J.L. McClelland and the PDP Research Group (1986). *Parallel Distributed Processing (vols 1 & 2)*
 - Nice simulators: Rochester, Stuttgart
 - Special hardware
- Applications in natural language processing
 - Parsing Complex Sentences with Structured Connectionist Networks
Jain, Neural Computing, 1991
- Applications in speech recognition
 - Phoneme Recognition Using Time-Delay Neural Networks
Waibel, Hanazawa, Hinton, Shikano, Lang
IEEE Transactions on Acoustics, Speech, and Signal Processing, 1989
- Applications in machine translation
 - Machine Translation using Neural Networks and Finite-State Models
Castaño, Casacuberta, Vidal, TMI 1997



Paradigm Shifts in Machine Translation



More Wind of Change – MT Business

- Translation has grown to a 40 billion USD industry
 - Yearly grows: 7%
- Machine translation is taking an increasing share
 - Yearly growth 20-25% pa *
- Observations:
 - Companies are embracing MT
 - Governments are adopting MT (e.g. European Parliament)
 - Translators are (often unhappily) accepting the reality of MT
 - Translation memory providers provide MT APIs
- MT is not so much replacing human translation, but covering before uncovered territory

* [http://www.researchandmarkets.com/reports/3138713/globalmachinetranslation market20152019](http://www.researchandmarkets.com/reports/3138713/globalmachinetranslation%20market20152019)



MT for Web, Social Media, Online Shopping



143 Billion words a day across
100 language combinations



Would need 200,000
translators



facebook

2 billion text translations on
a daily basis

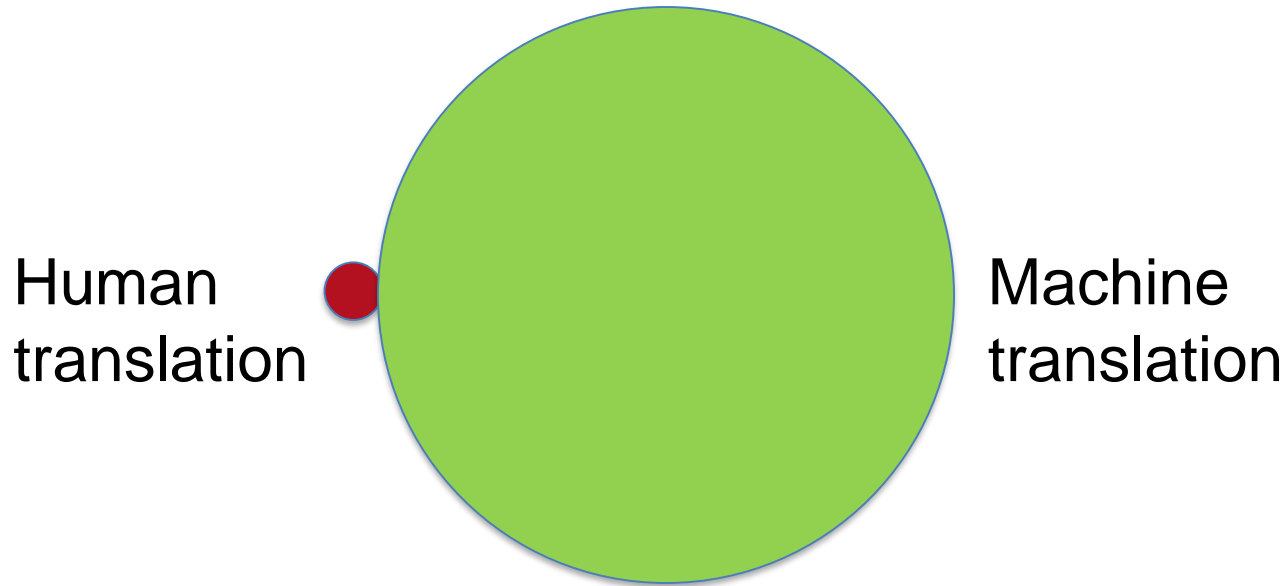


??? Billion words a day

- 1) <https://www.quora.com/How-many-words-are-translated-every-day-by-Google-translate>
- 2) <http://wersm.com/facebook-translates-billions-of-posts-every-day/>
- 3) Hassan Sawaf: Personal communication



MT for Web, Social Media, Online Shopping



100 times more content translated by machines than by humans!



Wind of Change - MT Startups

Disruption In Progress: 10 Years and 190 Translation Startups

by [Gino Diño](#) on October 6, 2015



The translation, localization, and language tech start-up scene is huge. In the past decade alone, 190 translation, language tech, and localization startups were founded, according to [Crunchbase's data](#).

<https://slator.com/features/disruption-in-progress-10-years-and-190-translation-startups/>



معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
عضو في مؤسسة قطر
Member of Qatar Foundation



مؤسسة قطر
Qatar Foundation

جامعة حمد بن خليفة

HAMAD BIN KHALIFA UNIVERSITY

Wind of Change - MT Startups

Some examples close to home:

- Mobile Technologies - acquired by Facebook

- Alex Waibel (KIT, CMU)
- Famous for Jibbigo



- Safaba – acquired by Amazon

- Alon Lavie (CMU)



- Apptek - MT section acquired by eBay

- Longwinded story. The short: Hassan Sawaf brings technology from RWTH to Apptek



- Unbabel: MT plus crowd-sourced postediting

- Former LTI PhD student: Vasco



- Lilt: interactive MT

- Out of Stanford: Spence Greene, John de Niro



Summary: MT and Winds of Change

- Another paradigm shift in MT research and development:
From SMT to NMT
- Machine translation is taking an increasing share in the translation industry

MT 20% growth versus HT 7% growth

- MT has become an everyday reality in the Web, in social media, etc
100 times more words by machine than by translators



Winds of Change - Exiting Times for MT Researchers



معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
عضو في مؤسسة قطر
Member of Qatar Foundation



مؤسسة قطر
Qatar Foundation

جامعة حمد بن خليفة

HAMAD BIN KHALIFA UNIVERSITY

Congratulations to InterACT & Alex



معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
عضو في مؤسسة قطر
Member of Qatar Foundation



مؤسسة قطر
Qatar Foundation

جامعة حمد بن خليفة

HAMAD BIN KHALIFA UNIVERSITY