2013 - word2vec

Tomas Mikolov et al. publish four papers on vector representations of words constituting the *word2vec* framework

This received very much attention as it revolutionized the way words were encoded for deep learning models in the field of NLP.

2013

2013 - word2vec

Tomas Mikolov et al. publish four papers on vector representations of words constituting the *word2vec* framework.

This received very much attention as it revolutionized the way words were encoded for deep learning models in the field of NLP.

2013 2014

2014 - Attention!

The concept of Attention is introduced by **Bahdanau et al.** as a complement to state-of-the-art neural network architectures.

It is (at first) just used in conjunction with RNNs for machine translation, but will kick-start a completely new path of Deep learning research in NLP in the years to come.

2013 - word2vec

Tomas Mikolov et al. publish four papers on vector representations of words constituting the *word2vec* framework

This received very much attention as it revolutionized the way words were encoded for deep learning models in the field of NLP.

2016 - TPUs

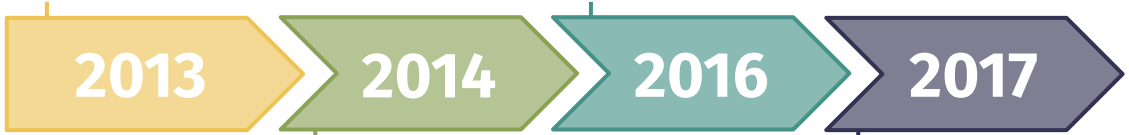
Google developed a processing unit exceptionally suited for deep learning applications.

Its architecture is optimized for performing large matrix multiplications and perfectly integrated to their deep learning library *Tensorflow*.

2014 - Attention!

The concept of Attention is introduced by **Bahdanau et al.** as a complement to state-of-the-art neural network architectures.

It is (at first) just used in conjunction with RNNs for machine translation, but will kick-start a completely new path of Deep learning research in NLP in the years to come.



2013 - word2vec

Tomas Mikolov et al. publish four papers on vector representations of words constituting the *word2vec* framework

This received very much attention as it revolutionized the way words were encoded for deep learning models in the field of NLP.

2016 - TPUs

Google developed a processing unit exceptionally suited for deep learning applications.

Its architecture is optimized for performing large matrix multiplications and perfectly integrated to their deep learning library *Tensorflow*.

2014 - Attention!

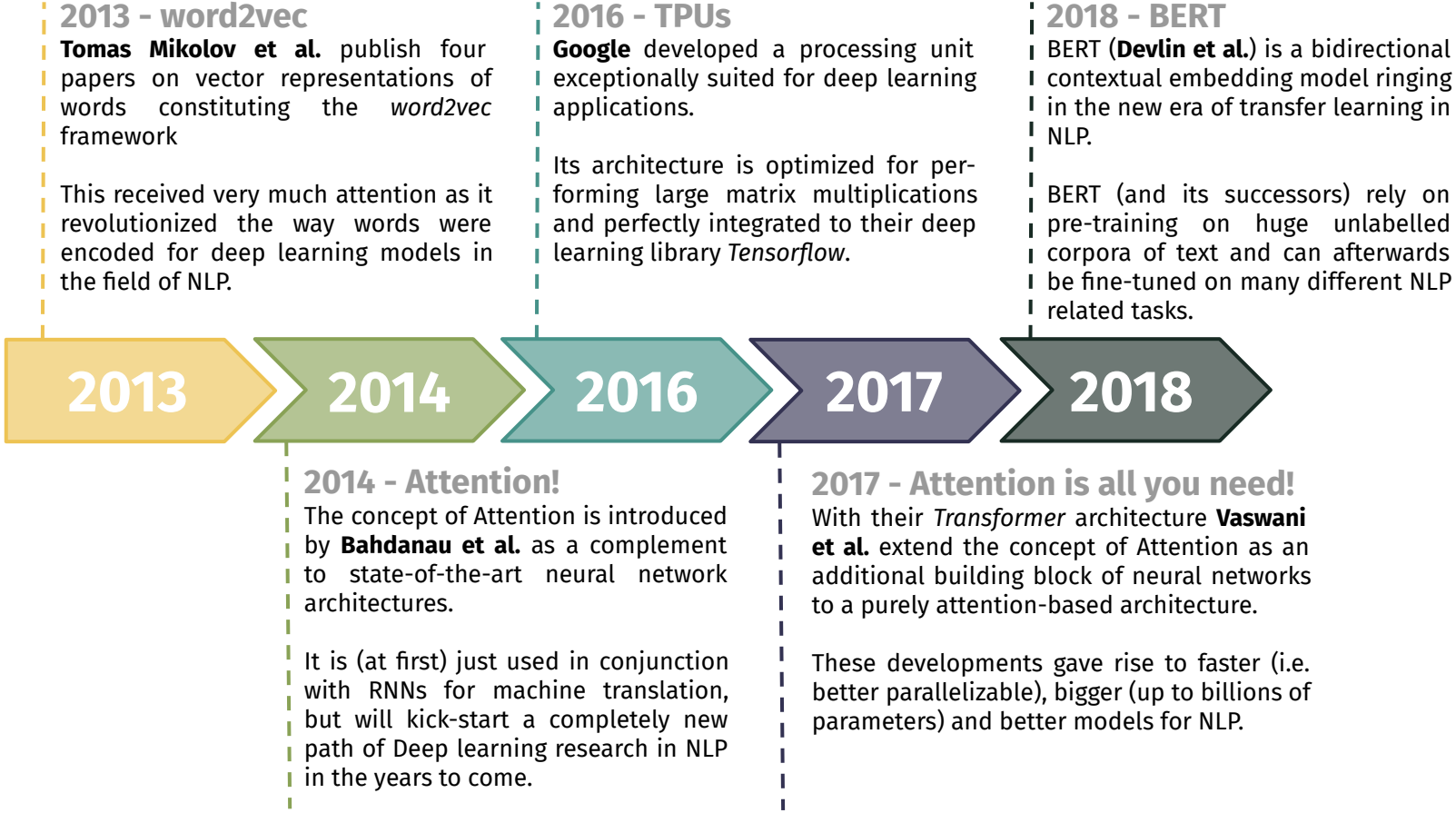
The concept of Attention is introduced by **Bahdanau et al.** as a complement to state-of-the-art neural network architectures.

It is (at first) just used in conjunction with RNNs for machine translation, but will kick-start a completely new path of Deep learning research in NLP in the years to come.

2017 - Attention is all you need!

With their *Transformer* architecture **Vaswani et al.** extend the concept of Attention as an additional building block of neural networks to a purely attention-based architecture.

These developments gave rise to faster (i.e. better parallelizable), bigger (up to billions of parameters) and better models for NLP.


2013 - word2vec
Tomas Mikolov et al. publish four papers on vector representations of words constituting the word2vec framework
This received very much attention as it revolutionized the way words were encoded for deep learning models in the field of NLP.
2016 - TPUs
Google developed a processing unit exceptionally suited for deep learning applications.
Its architecture is optimized for performing large matrix multiplications and perfectly integrated to their deep learning library Tensorflow.
2018 - BERT
BERT (Devlin et al.) is a bidirectional contextual embedding model ringing in the new era of transfer learning in NLP.
BERT (and its successors) rely on pre-training on huge unlabelled corpora of text and can afterwards be fine-tuned on many different NLP related tasks.
2013
2014
2016
2017
2018
2014 - Attention!
The concept of Attention is introduced by Bahdanau et al. as a complement to state-of-the-art neural network architectures.
It is (at first) just used in conjunction with RNNs for machine translation, but will kick-start a completely new path of Deep learning research in NLP in the years to come.
2017 - Attention is all you need!
With their Transformer architecture Vaswani et al. extend the concept of Attention as an additional building block of neural networks to a purely attention-based architecture.
These developments gave rise to faster (i.e. better parallelizable), bigger (up to billions of parameters) and better models for NLP.