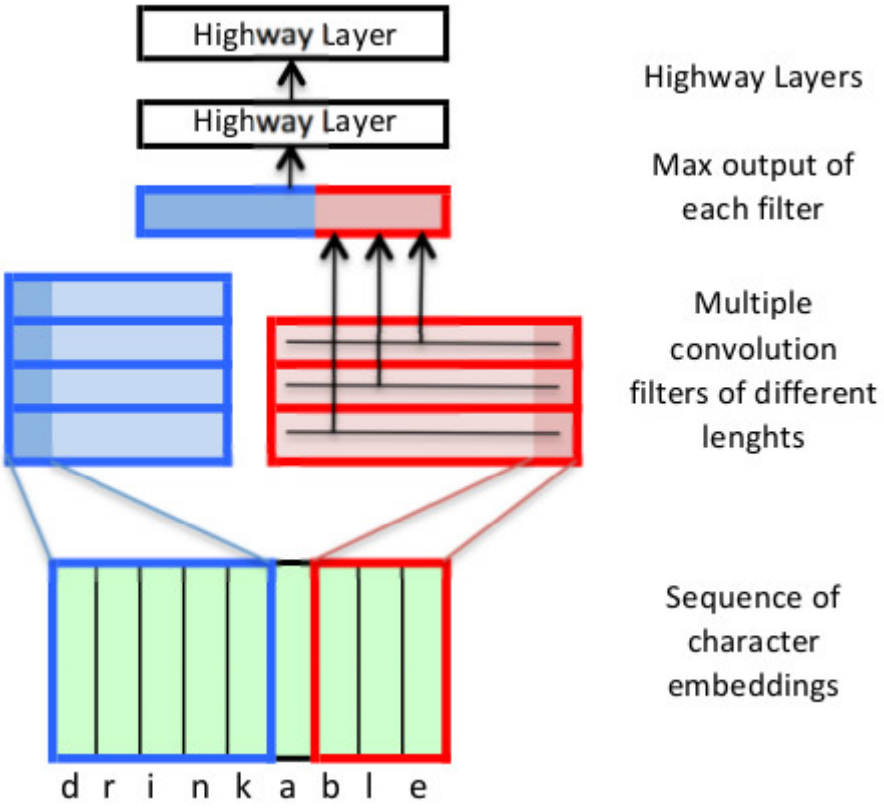
Paper

- **Title**: Character-based Neural Machine Translation
- **Authors**: Marta R. Costa-Jussà, José A. R. Fonollosa
- **Link**: http://arxiv.org/abs/1603.00810v3
- **Tags**: Neural Network, machine translation
- **Year**: 2016

Summary

- What
  - Most neural machine translation models currently operate on word vectors or one hot vectors of words.
  - They instead generate the vector of each word on a character-level.
    - Thereby, the model can spot character-similarities between words and treat them in a similar way.
    - They do that only for the source language, not for the target language.
- How
  - They treat each word of the source text on its own.

    - To each word they then apply the model from Character-aware neural language models, i.e. they do per word:
        - Embed each character into a 620-dimensional space.
        - Stack these vectors next to each other, resulting in a 2d-tensor in which each column is one of the vectors (i.e. shape `620xN` for `N` characters).
        - Apply convolutions of size `620xW` to that tensor, where a few different values are used for `W` (i.e. some convolutions cover few characters, some cover many characters).
        - Apply a tanh after these convolutions.
        - Apply a max-over-time to the results of the convolutions, i.e. for each convolution use only the maximum value.
        - Reshape to 1d-vector.
        - Apply two highway-layers.
        - They get 1024-dimensional vectors (one per word).
        - Visualization of their steps:
        - 

    - Afterwards they apply the model from Neural Machine Translation by Jointly Learning to Align and Translate to these vectors, yielding a translation to a target language.
    - Whenever that translation yields an unknown target-language-word ("UNK"), they replace it with the respective (untranslated) word from the source text.

- Results
    - They the German-English WMT dataset.

- BLEU improvemements (compared to neural translation without character-level words):
  - German-English improves by about 1.5 points.
  - English-German improves by about 3 points.
- Reduction in the number of unknown target-language-words (same baseline again):
  - German-English goes down from about 1500 to about 1250.
  - English-German goes down from about 3150 to about 2650.
- Translation examples (Phrase = phrase-based/non-neural translation, NN = non-character-based neural translation, CHAR = theirs):

1	SRC	Berichten zufolge hofft Indien darber hinaus auf einen Vertrag zur **Verteidigungszusammenarbeit** zwischen den beiden Nationen .
	Phrase	reportedly hopes India , in addition to a contract for the defence cooperation between the two nations .
	NN	according to reports , India also hopes to establish a contract for the UNK between the two nations .
	CHAR	according to reports , India hopes to see a Treaty of **Defence Cooperation** between the two nations .
	REF	India is also reportedly hoping for a deal on **defence collaboration** between the two nations .
2	SRC	der durchtrainierte Mainzer sagt von sich , dass er ein " ambitionierter **Rennradler** " ist .
	Phrase	the will of Mainz says that he a more ambitious .
	NN	the UNK Mainz says that he is a " ambitious , . "
	CHAR	the UNK in Mainz says that he is a ' **ambitious racer** ' .
	REF	the well-conditioned man from Mainz said he was an " **ambitious racing cyclist** . "
3	SRC	die GDL habe jedoch nicht gesagt , wo sie **streiken** wolle , so dass es schwer sei , die Folgen konkret vorherzusehen .
	Phrase	the GDL have , however , not to say , where they strike , so that it is difficult to predict the consequences of concrete .
	NN	however , the UNK did not tell which they wanted to UNK , so it is difficult to predict the consequences .
	CHAR	however , the UNK did not say where they wanted to **strike** , so it is difficult to predict the consequences .
	REF	the GDL have not said , however , where they will **strike** , making it difficult to predict exactly what the consequences will be .
4	SRC	die Premierminister Indiens und Japans trafen sich in Tokio .
	Phrase	the Prime Minister of India and Japan in Tokyo .
	NN	the Prime Minister of India and Japan met in Tokyo
	CHAR	the Prime **Ministers** of India and Japan met in Tokyo
	REF	India and Japan prime **ministers** meet in Tokyo
5	SRC	wo die Beamten es aus den Augen verloren .
	Phrase	where the officials lost sight of
	NN	where the officials lost it out of the eyes
	CHAR	where officials **lose sight of it**
	REF	causing the officers to **lose sight of it**