

Natural Language Processing (FS 2022)

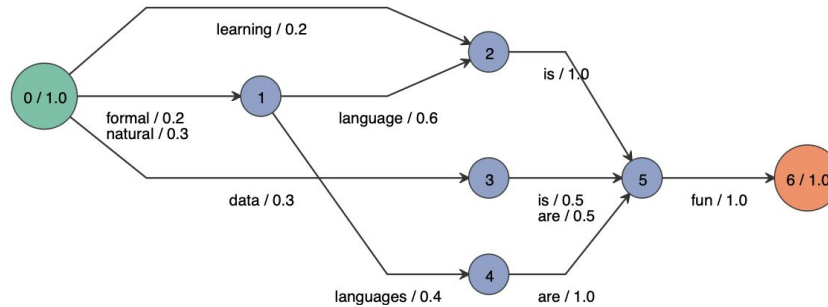
Transliteration with WFSTs

09.11.2022

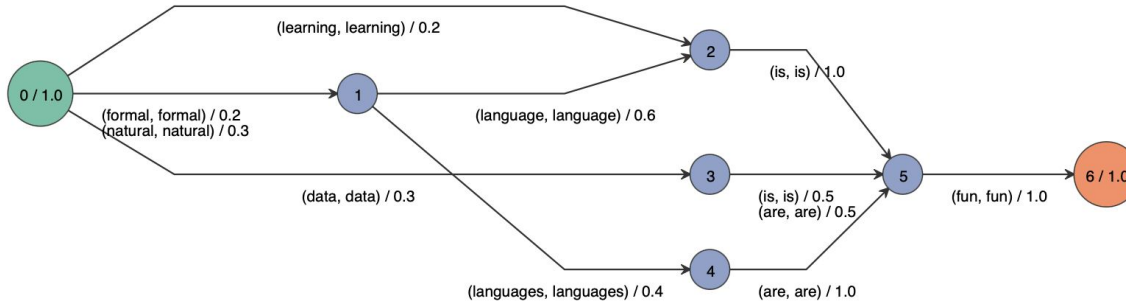
1. WFSA to WFST

- Any WFSA can be encoded as a WFST

- WFSA:



- WFST:



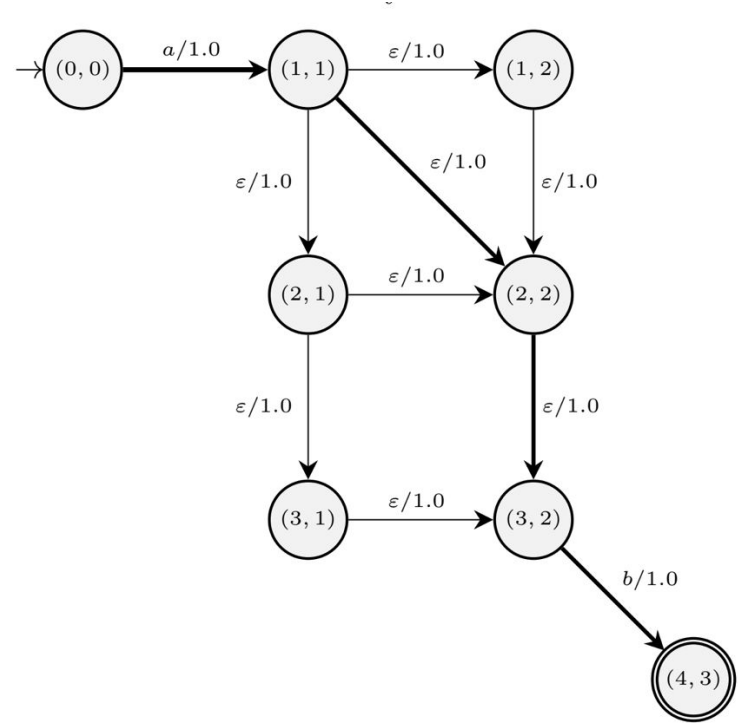
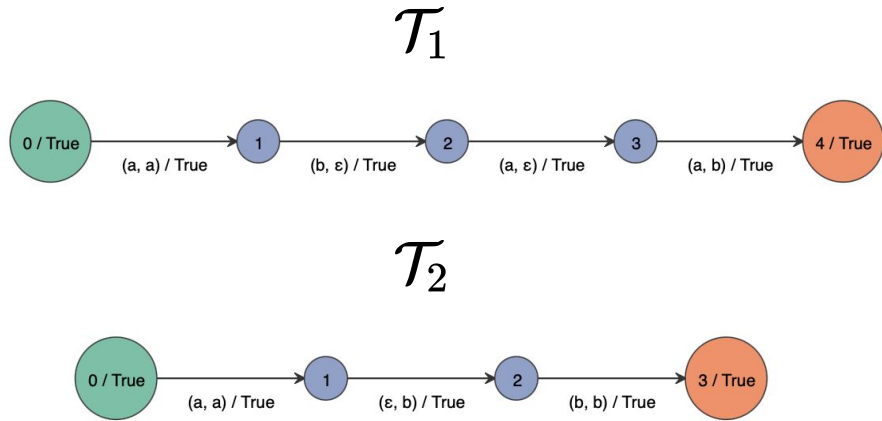
2. Transducer Composition

- Given two transducers \mathcal{T}_1 and \mathcal{T}_2 we can construct the composition \mathcal{T}

$$\mathcal{T}(\mathbf{x}, \mathbf{y}) = \bigoplus_{\mathbf{z} \in \Omega^*} \mathcal{T}_1(\mathbf{x}, \mathbf{z}) \otimes \mathcal{T}_2(\mathbf{z}, \mathbf{y})$$

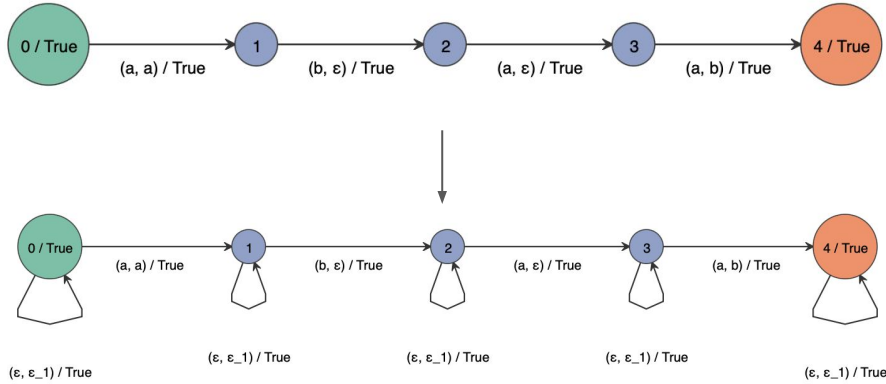
- For each \mathbf{z} , we must match one path in \mathcal{T}_1 with output \mathbf{z} and one path in \mathcal{T}_2 with input \mathbf{z}
- Standard algorithms fail for \mathcal{E} -transitions

2. Transducer Composition

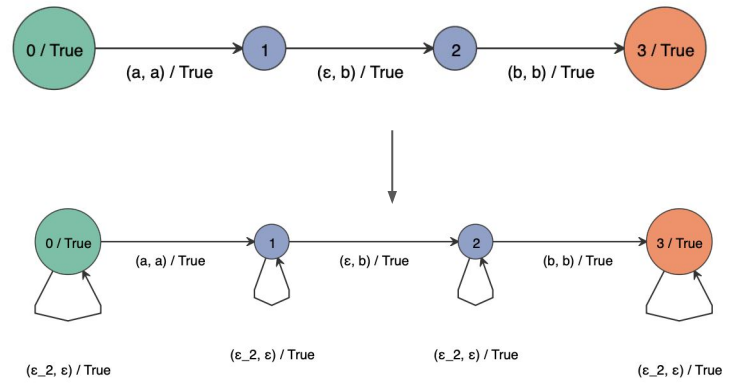


2. Augmentation

\mathcal{T}_1

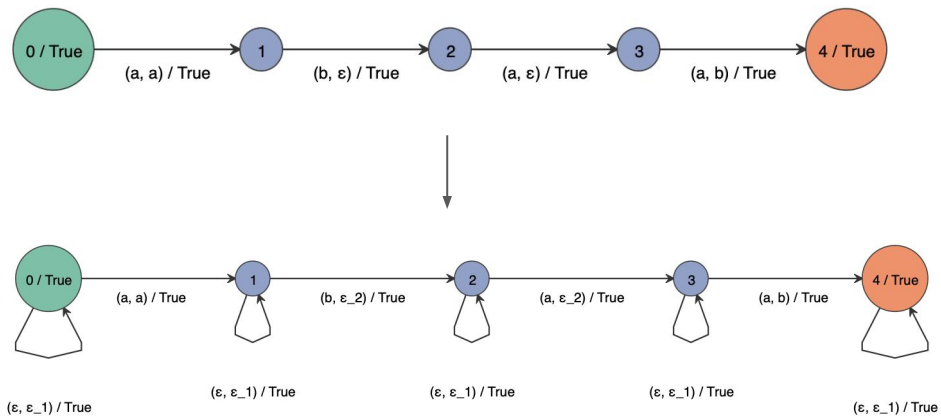


\mathcal{T}_2

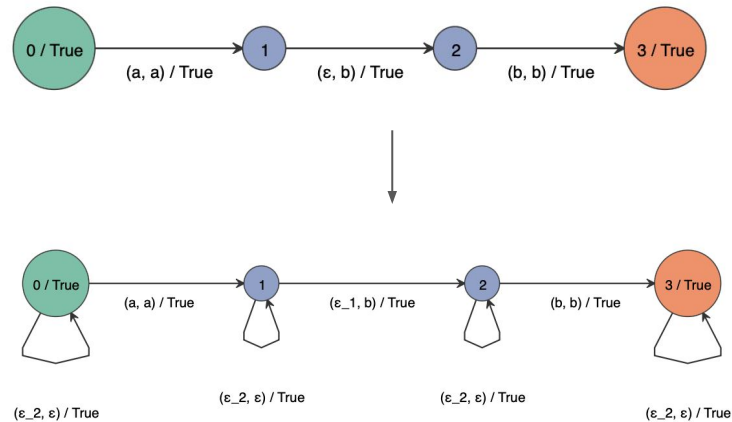


2. Relabeling

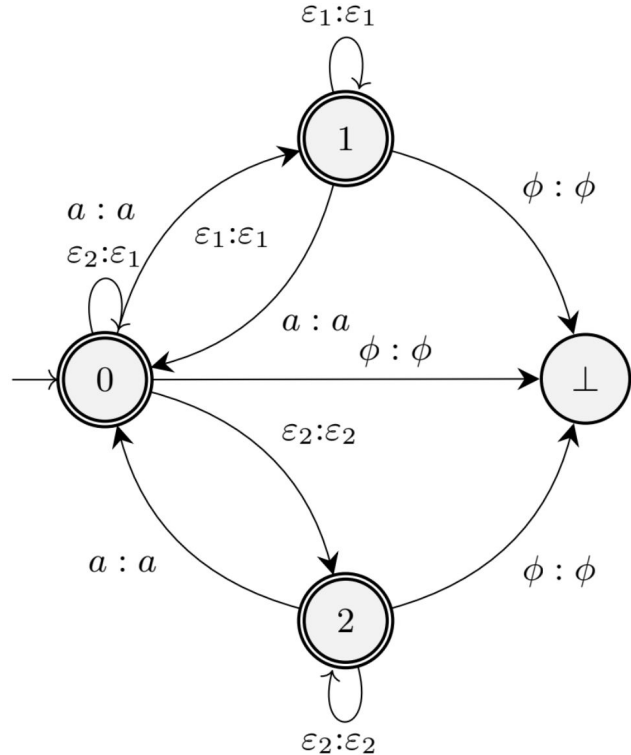
\mathcal{T}_1



\mathcal{T}_2



2. ε -Filter



- Each state keeps track of what actions have been performed previously.
- The blocking state ensures that only one path is allowed.
- a is any symbol in the alphabet.
- $\phi : \phi$ is anything that is not defined.