D is a set of documents. We will suppose each document has one single sentance for now.

Let d be a document in D.

For d is a set of sequence of tokens $(t_i)_{0 \leq i < n}$

We associate embeddings $(u_i)_{0 \leq i < n}$

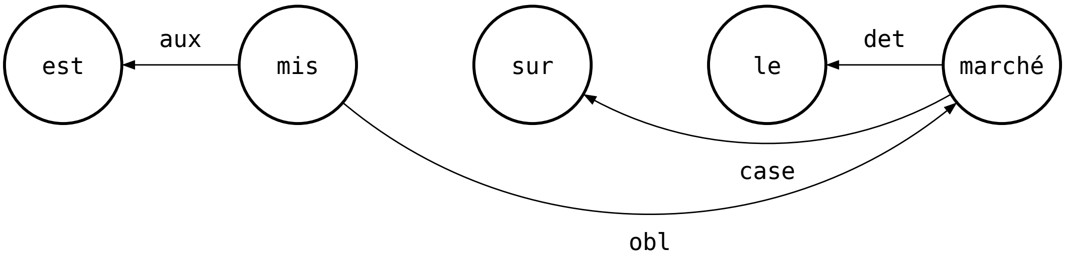
And a tree structure represented by it's adjacency matrix T. T_{ij} means that i if the father of j in the tree, according to dependency parsing.

Fome some tree T, we can multiply it by a diagonal matrix with boolean coefficients, C. We will refer to this matrix a the "cut".

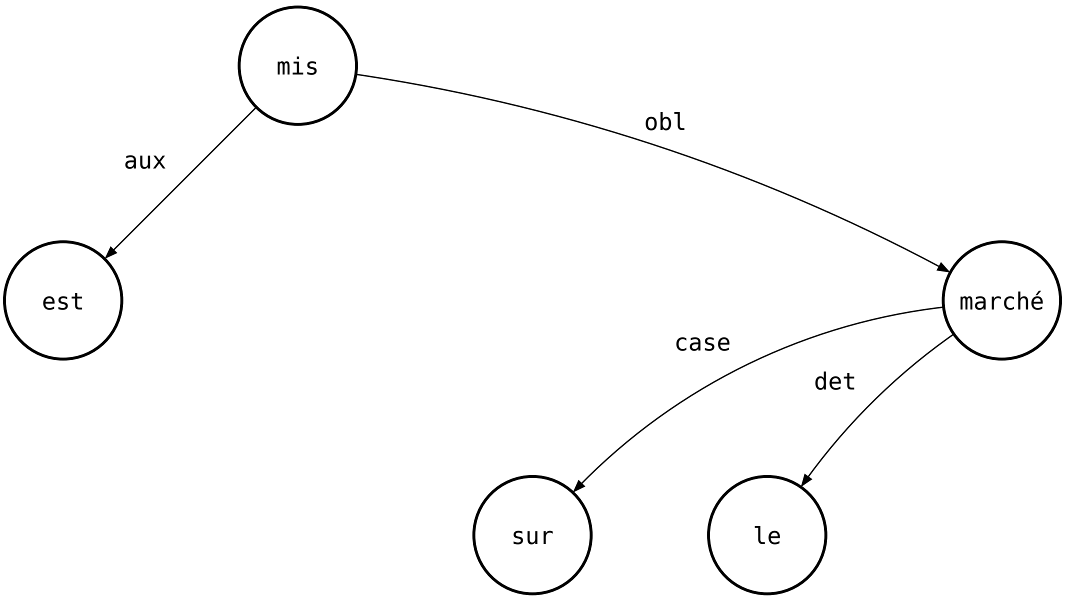
We will use the following notation to describe this bilinear map:

$$\Phi(T, C) = T \ \mathrm{diag}(C_1, C_2, ..., C_n)$$

For example, let's take the following document:



This is more classically represented as:



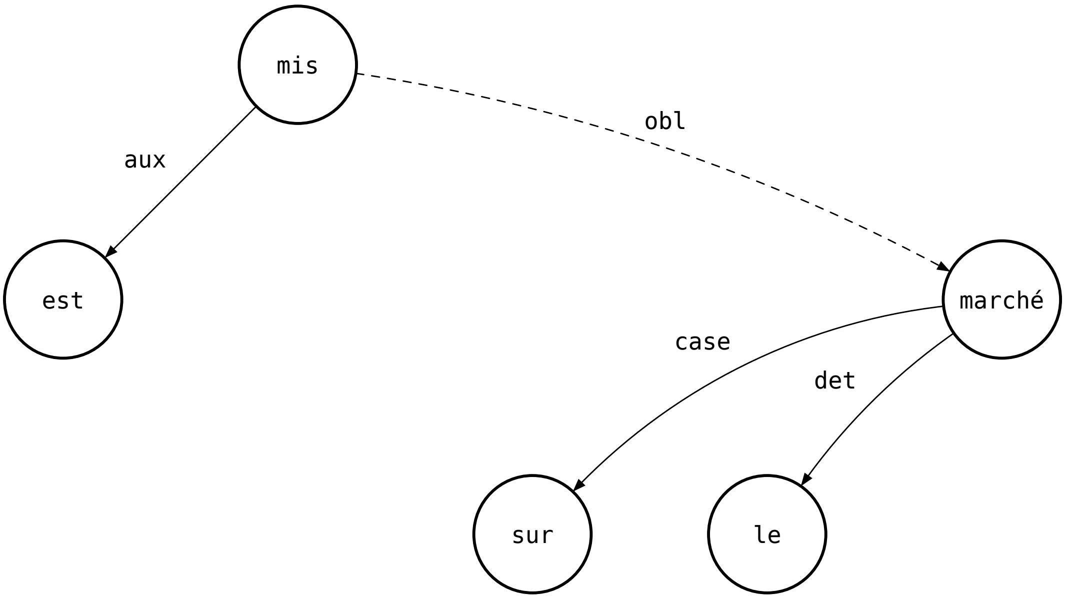
The associated matrix is

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Using the cut $C = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$, he have

$$\Phi(T, C) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

We get a new graph with the edge between the third word and its parent removed.



We get 2 subtrees. In general, we get the new roots of the subtree by considering $\mathrm{Ker}(TC)$

We not $\mathrm{sub}(T)$ the matrix containing on each line i, the set of descendents of the node i[1]

We can then express one single subtree with $\Phi(T, \mathrm{sub}(TC)E_i)$

Let's consider some score $s : \{0, 1\}^{n^2} \to \mathbb{R}$

The reward of some cut over a tree is: $R(T, C) = \sum_{i=1}^{n} s(\Phi(T, \mathrm{sub}(TC)E_i))$

We define a parametric model over the cuts of a tree:

$\pi_\theta(T) = \sum_{i=1}^{n} \frac{\exp(w_\theta \cdot u_i)}{Z_\theta}$

[1]There is a close form:

$$\begin{aligned} \mathrm{sub}(T) &= T^t + (T^2)t + \ldots + (T^n)^t \\ &= (I_n - T^t)^{-1} \end{aligned}$$