Tree-to-language correction (approximation)
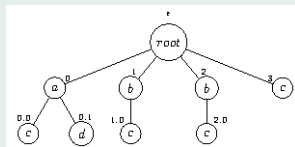
Invalid XML tree
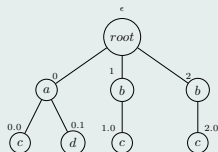
DTD

Closest valid XML tree

$root \rightarrow b^* | ab^* c$
$a \rightarrow cd$
$b \rightarrow c$
$c \rightarrow \epsilon$
$d \rightarrow \epsilon$

MWE identification in a treebank

- MWE = a tree or a **tree family** (e.g. as in LTAG),

- occurrence of a MWE in a treebank = a **subtree**,

- MWE identification = correcting the **subtree** wrt. the **tree family**,

- advantage: robust identification in **noisy data**.