$\mathbb{R}$	The set of real numbers.
$\mathbf{a}, \mathbf{b}, \ldots$	Vectors over the reals, *i.e.* $\mathbf{a} \in \mathbb{R}^m$.
$\mathbf{A}, \mathbf{B}, \ldots$	Matrices over the reals, *i.e.* $\mathbf{A} \in \mathbb{R}^{m \times n}$.
$[\mathbf{a}]_i$	Vector indexing, $[\mathbf{a}]_i \in \mathbb{R}$ for $1 \leq i \leq m$.
$[\mathbf{a}; \mathbf{b}]$	Vector concatenation, $\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^n$, $[\mathbf{a}; \mathbf{b}] \in \mathbb{R}^{m+n}$.
$[\mathbf{a}, \mathbf{b}]$	Vertical vector stacking, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$, $[\mathbf{a}, \mathbf{b}] \in \mathbb{R}^{m \times 2}$.
$\mathcal{X}$	Finite vocabulary of words x.
$\mathcal{Y}(x)$	Finite set of trees y that are compatible with x.
$X, Y, \ldots$	Random variables with sample spaces $\mathcal{X}, \mathcal{Y}, \ldots$
x	A word from vocabulary $\mathcal{X}$, an outcome of random variable X.
y	A tree from $\mathcal{Y}(x)$, an outcome of random variable Y.
x_1^m, x	A sequence of words $\langle x_1, \ldots, x_m \rangle$ from $\mathcal{X}^m$ (shorthand: x), the outcome of the sequence of random variables $X_1, X_2, \ldots, X_m$.
$x_{<i}$	The sequence x_1^{i-1}, the part of sequence x_1^m preceding word x_i.
P_X, $P_{Y\|X}$, $P_{X,Y}$	Discrete probability distributions of random variables X, conditional $X\|Y$ and joint X, Y
p_X, $p_{Y\|X}$, $p_{Y\|X}$	Probability mass functions of distributions P_X, $P_{Y\|X}$, and $P_{X,Y}$.
$p(x)$, $p(y \mid x)$, $p(x, y)$	Shorthands for probabilities $P_X(X = x)$, $P_{X\|Y}(Y = y \mid X = x)$, and $P_{X,Y}(X = x, Y = y)$.
p_θ, q_λ	Probability mass functions with emphasis on the parameters that specify them, and with random variables clear from context.
$\mathbb{E}[g(X)]$	Expectation of $g(X)$ with respect to distribution P_X, for some real-valued function g.
$H(P_X)$	Entropy of random variable X with distribution P
Λ	Finite set of nonterminal labels in a context-free grammar.
$A, B, C, \ldots$	Nonterminal labels from Λ.
$S^\dagger$	Special root label not in Λ.