Bezkontextové grafové gramatiky

Zdeněk Sawa

8. října 1999

1 Úvod

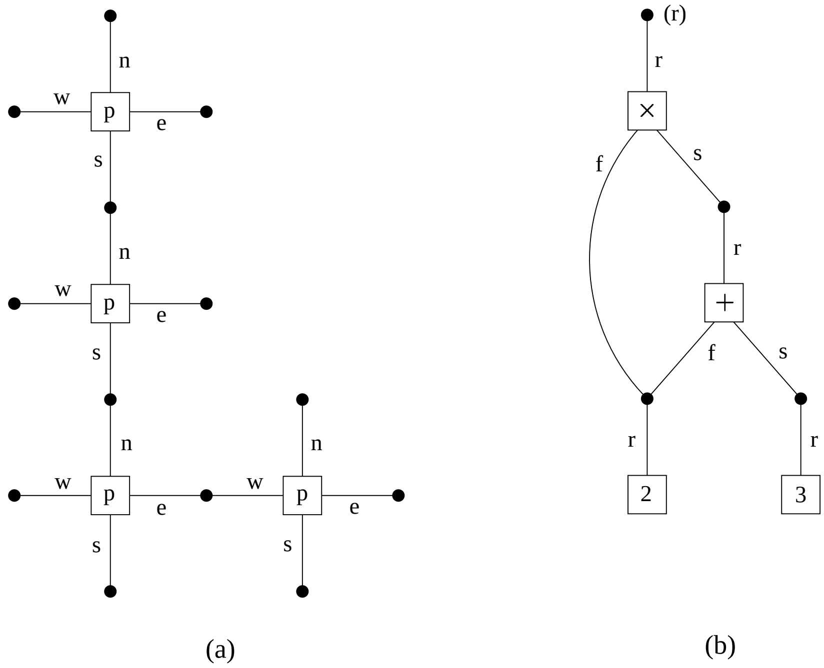
Pojem jazyk tak jak ho chápeme v teorii formálních jazyků, to znamená jako množinu slov nad danou abecedou, můžeme použít rovněž v teorii grafů, kde pod pojmem ***grafový jazyk*** budeme rozumět množinu grafů. Jedním z možných způsobů jak grafové jazyky popsat jsou tzv. ***grafové gramatiky***.

Pro celou řadu grafových jazyků platí, že grafy patřící do daného jazyka mají jakoby stromovou strukturu v tom smyslu, že mohou být rekurzivně sestaveny z určitých základních grafů pomocí určitých základních operací. Takovéto grafové jazyky mohou být generovány pomocí ***bezkontextových grafových gramatik***, což jsou systémy pro přepisování grafů, ve kterých použití přepisovacího (produkčního) pravidla spočívá v odstranění jedné atomické části grafu (tj. jednoho vrcholu nebo jedné hrany) a nahrazení této atomické části grafem, který je připojen k původnímu grafu na její místo. Použití přepisovacího pravidla by mělo být nezávislé na kontextu, ve kterém se nachází nahrazovaný vrchol nebo hrana, tj. na ostatních vrcholech a hranách původního grafu. V tomto smyslu jsou bezkontextové grafové gramatiky přirozeným zobecněním „klasických“ bezkontextových gramatik, které generují řetězce symbolů.

Ovšem na rozdíl od bezkontextových gramatik, které generují řetězce symbolů, je možné definovat několik různých typů bezkontextových grafových gramatik. Mohou se lišit například v následujících rysech:

- typ grafů, které jsou těmito gramatikami generovány (např. orientované grafy, neorientované grafy, hypergrafy apod.),
- atomické části grafů, které jsou nahrazovány při použití přepisovacích pravidel (tj. vrcholy nebo hrany),
- jakým způsobem je graf, kterým nahradíme vrchol nebo hranu, připojen k původnímu grafu (např. ztotožněním některých vrcholů nebo přidáním nových hran).

Při zkoumání různých typů těchto gramatik se jako nejpřirozenější a nejvhodnější ukázaly dva typy gramatik: HR gramatiky (hyperedge replacement grammars) a C-edNCE gramatiky (context-free graph grammars with neighbourhood controlled embedding and dynamic edge relabeling) neboli NR gramatiky (node replacement grammars). Tyto dva typy bezkontextových grafových gramatik budou popsány v následujícím textu.



Obrázek 1: Příklady hypergrafů

2 HR gramatiky

HR gramatiky (hyperedge replacement grammars) generují orientované hypergrafy s ohodnocenými hyperhranami. V těchto gramatikách jsou při použití přepisovacích pravidel nahrazovány hyperhrany hypergrafy, přičemž tyto hypergrafy jsou připojovány k původnímu hypergrafu ztotožněním některých vrcholů hypergrafů.

2.1 Hypergrafy

Hypergrafy (hypergraphs) jsou zobecněním „běžných" grafů. Stejně jako ony jsou tvořeny množinou vrcholů a množinou hran, kterým se ovšem říká hyperhrany. Zatímco v „běžných" grafech je každá hrana incidentní právě se dvěma vrcholy, v hypergrafech mohou být hyperhrany incidentní s libovolným počtem vrcholů.

Hypergrafy je možné znázornit graficky. V následujícím textu bude používána následující konvence: vrcholy budou znázorněny jako body a hyperhrany jako čtverce. Čáry, které spojují hrany s incidentními vrcholy se nazývají ***tykadla*** (tentacles). Každé takovéto tykadlo je označeno tzv. ***selektorem*** (selector), který vyznačuje vztah mezi hranou a vrcholem. Různá tykadla jedné hrany musí být ohodnocena různými selektory. ***Typ hrany*** (type of an edge) je množina selektorů, které ohodnocují její tykadla. Hrany v hypergrafu nemusí být všechny stejného typu.

Pozn.: V následujícím textu se kvůli stručnosti někdy používá místo termínu hypergraf pojem graf a místo termínu hyperhrana pojem hrana.

Dva příklady hypergrafů jsou na obr. 1. Na obr. 1(a) je znázorněn hypergraf, který má 4 hrany a 13 vrcholů, přičemž každá hrana je incidentní se čtyřmi vrcholy. Na obr. 1(b) je znázorněn hypergraf, který vyjadřuje aritmetický výraz $2 \times (2 + 3)$.

Abychom byli schopni hypergrafy spojovat dohromady, označujeme některé vrcholy hypergrafu jako ***externí*** (external nodes). Tyto vrcholy představují určité „rozhraní“ hypergrafu, pomocí něhož může být tento hypergraf připojen k jinému hypergrafu. Tyto externí vrcholy jsou ohodnoceny pomocí selektorů. Při grafickém znázorňování hypergrafů jsou tyto vrcholy ohodnoceny selektorem v závorkách (např. na obr. 1(b) je to vrchol ohodnocený (r), přičemž r je příslušný selektor). Význam těchto externích vrcholů bude vysvětlen dále.

Následuje formální definice hypergrafu.

Definice 2.1 ***Hypergraf*** nad Γ a Σ je pětice $H = (V, E, \mathrm{lab}, \mathrm{nod}, \mathrm{ext})$, kde

Γ — je abeceda symbolů, kterými mohou být ohodnoceny hyperhrany,
Σ — je množina (abeceda) selektorů,
V — je konečná množina vrcholů,
E — je konečná množina hyperhran,
lab — je funkce $E \to \Gamma$, která slouží k ohodnocení hyperhran, každé hyperhraně je přiřazen symbol z množiny Γ,
nod — je incidenční funkce přiřazující každé hyperhraně vrcholy incidentní s touto hranou, každé hyperhraně $e \in E$ je přizazena parciální funkce $\mathrm{nod}(e) : \Sigma \to V$,
ext — je externí funkce, je to parciální funkce $\Sigma \to V$.

Typ grafu je množina selektorů, kterými jsou ohodnoceny externí vrcholy grafu. Tuto množinu označujeme $\mathrm{type}(H)$, přičemž $\mathrm{type}(H) = \mathrm{dom}(\mathrm{ext})$. ***Typ hrany*** je množina selektorů, kterými jsou ohodnocena tykadla dané hyperhrany. Tuto množinu označujeme $\mathrm{type}(e)$, přičemž pro libovolnou hranu e platí, že $\mathrm{type}(e) = \mathrm{dom}(\mathrm{nod}(e))$. □

Pozn.: Zápisem $\mathrm{dom}(f)$ označujeme definiční obor funkce f, tj. množina prvků pro něž je hodnota funkce f definována.

Jednotlivé komponenty grafu H můžeme též označovat V_H, E_H, lab_H, nod_H a ext_H. Oborem hodnot funkce $\mathrm{nod}(e)$ je množina vrcholů incidentních s hranou e. Zápisem $\mathrm{nod}(e)(s)$ označujeme vrchol incidentní s hranou e určený selektorem s. Místo zápisu $\mathrm{nod}(e)(s)$ je též možné používat přehlednější zápis $\mathrm{nod}(e, s)$.

Dva hypergrafy jsou ***isomorfní*** (isomorphic) pokud se liší pouze identitou vrcholů a hran. To znamená, že hypergrafy H a K jsou isomorfní právě když existují bijektivní zobrazení $g : V_H \to V_K$ a $h : E_H \to E_K$ taková, že pro všechna $e \in E_H$ a $s \in \Sigma$ platí, že $\mathrm{lab}_K(h(e)) = \mathrm{lab}_H(e)$, $\mathrm{nod}_K(h(e), s) = g(\mathrm{nod}_H(e, s))$ a $\mathrm{ext}_K(s) = g(\mathrm{ext}_H(s))$. Dva grafy jsou ***vrcholově isomorfní*** (node-isomorphic) pokud se liší pouze identitou svých vrcholů. To znamená, že pro dva isomorfní grafy H a K platí, že jsou vrcholově isomorfní právě když $E_H = E_K$. V následujícím textu budou vrcholově isomorfní grafy považovány za jeden a tentýž graf, což budeme označovat zápisem $H = K$.

Dva hypergrafy H a K jsou ***disjunktní*** (disjoint) jestliže platí, že $E_H \cap E_K = \emptyset$, tj. pokud jsou disjunktní množiny jejich hran. Vzhledem k tomu, že vrcholově isomorfní grafy považujeme za jeden a tentýž graf, můžeme navíc předpokládat, že $(V_H \cup E_H) \cap (V_K \cup E_K) = \emptyset$, tj. že jsou disjunktní i množiny vrcholů (pokud nejsou, můžeme vrcholy nahradit jinými, přičemž graf se tím nezmění).

Nechť H je hypergraf, pak množinu všech grafů isomorfních s tímto grafem označujeme $[H]$. Na množinu $[H]$ se můžeme dívat jako na „abstraktní“ hypergraf, zatímco H je „konkrétní“ hypergraf, instance abstraktního hypergrafu $[H]$.

Množinu všech (konkrétních) hypergrafů nad Γ a Σ označujeme $HGR_{\Gamma,\Sigma}$ a množinu všech abstraktních hypergrafů $[HGR_{\Gamma,\Sigma}]$. ***Hypergrafový jazyk*** L je podmnožina množiny $[HGR_{\Gamma,\Sigma}]$, přičemž platí, že všechny hypergrafy patřící do množiny L musí být stejného typu.

2.2 Reprezentace „obyčejných“ orientovaných grafů pomocí hypergrafů

Pokud pro hyperhranu e platí, že $\mathrm{type}(e) = \{\mathrm{sou}, \mathrm{tar}\}$, nazýváme tuto hranu ***obyčejná hrana*** (ordinary edge). Na obrázcích je možné nakreslit tuto hranu jako šipku vedoucí z vrcholu určeného selektorem sou do vrcholu určeného selektorem tar, přičemž tato hrana je ohodnocena symbolem $\mathrm{lab}(e)$. Hypergraf, který obsahuje pouze obyčejné hrany odpovídá „obyčejnému“ orientovanému grafu s ohodnocenými hranami.

Dalším speciálním typem hyperhrany je tzv. ***značka*** (tag), pro kterou platí, že $\mathrm{type}(e) = \{\mathrm{owner}\}$. Hrany tohoto typu je možno použít k ohodnocení vrcholů, protože symbol jímž je ohodnocena (tj. $\mathrm{lab}(e)$) můžeme považovat za ohodnocení (jediného) vrcholu s nímž je tato hrana incidentní. Na obrázcích je možné vrchol a značku zobrazit jako ohodnocený vrchol.

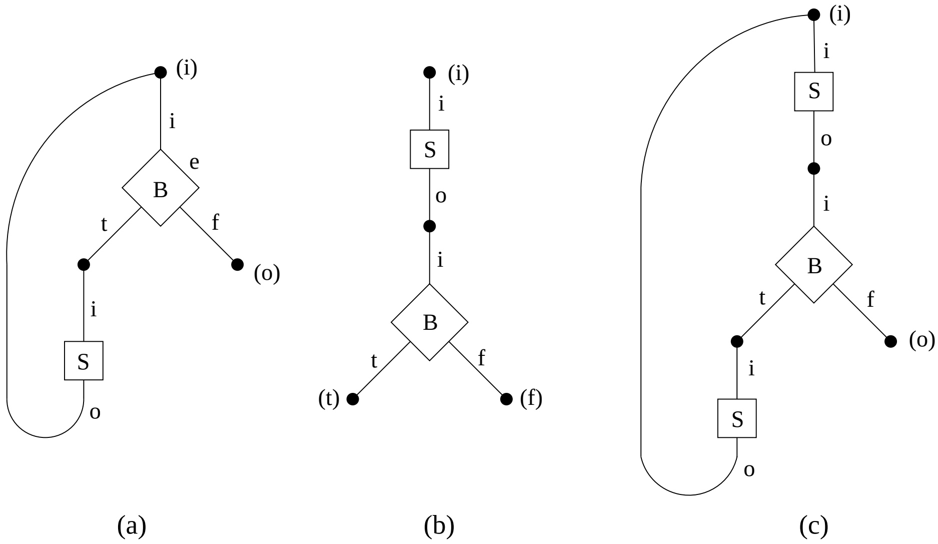
2.3 Náhrada hyperhran

V HR gramatikách jsou hyperhrany ohodnoceny terminálními a neterminálními symboly (terminály a neterminály). Přepisovací pravidla jsou typu $X \to D$, kde X je neterminální symbol a D je hypergraf s externími vrcholy. Toto přepisovací pravidlo můžeme použít k ***náhradě*** neboli substituci (substitution) hrany e grafu H grafem D, přičemž musí platit, že hrana e je ohodnocena symbolem X, tj. $\mathrm{lab}(e) = X$, a hrana e i graf K jsou stejného typu, tj. $\mathrm{type}(e) = \mathrm{type}(K)$. Použití pravidla $X \to D$ spočívá v odstranění hrany e z grafu H (spolu s jejími tykadly), přidání isomorfní kopie grafu K ke zbytku grafu H (grafy H a K musí být disjunktní) a ve ztotožnění vrcholů, které byly v grafu H incidentní s hranou e, s externími vrcholy grafu K, přičemž ztotožňujeme vždy vrcholy označené týmiž selektory. Výsledný graf, který vznikne náhradou hrany e v grafu H grafem K označujeme zápisem $H[e/K]$.

Ztotožnění vrcholů je definováno takto: Nechť H je hypergraf a $R \subseteq V_H \times V_H$ je relace, která obsahuje dvojice vrcholů, které chceme ztotožnit (tj. pokud chceme ztotožnit v a v', pak $(v, v') \in R$). Symbolem $\equiv_R$ označme nejmenší relaci typu ekvivalence (nad množinou vrcholů V_H) obsahující relaci R. Pak třídu rozkladu do které patří vrchol $v \in V_H$ v ekvivalenci $\equiv_R$ označíme $[v]_R$ a množinu těchto tříd rozkladu označíme $V_H/{\equiv_R}$, tj. $V_H/{\equiv_R} = \{[v]_R : v \in V_H\}$. Graf, který získáme ztotožněním vrcholů podle relace R (budeme ho označovat H/R), bude definován takto: $H/R = (V_H/{\equiv_R}, E_H, \mathrm{lab}_H, \mathrm{nod}, \mathrm{ext})$, kde pro každou hranu e a každý selektor s platí, že $\mathrm{nod}(e, s) = [\mathrm{nod}_H(e, s)]_R$ a $\mathrm{ext}(s) = [\mathrm{ext}_H(s)]_R$, tj. vrcholy nahradíme třídami rozkladu podle ekvivalence $\equiv_R$.

Definice 2.2 Nechť $H, K \in HGR_{\Gamma,\Sigma}$ jsou disjunktní hypergrafy a nechť $e \in E_H$ je hrana stejného typu jako hypergraf K, tj. $\mathrm{type}(e) = \mathrm{type}(K)$. Zápisem $(H - e) + K$ označujeme hypergraf, který dostaneme po odstranění hrany e z grafu H a po přidání hypergrafu K, tj. $(H - e) + K = (V, E, \mathrm{lab}, \mathrm{nod}, \mathrm{ext})$, kde $V = V_H \cup V_K$, $E = (E_H - \{e\}) \cup E_K$, lab je zúžení (restrikce) $\mathrm{lab}_H \cup \mathrm{lab}_K$ na množinu E, nod je zúžení $\mathrm{nod}_H \cup \mathrm{nod}_K$ na množinu E a $\mathrm{ext} = \mathrm{ext}_H$. Náhradou (substitucí) hrany e v grafu H grafem K dostaneme hypergraf $H[e/K] = ((H - e) + K)/R$, kde $R = \{(\mathrm{nod}_H(e, s), \mathrm{ext}_K(s)) : s \in \mathrm{type}(e)\}$. $\square$

Dá se snadno ověřit, že pokud H a H' jsou vrcholově isomorfní a K a K' jsou vrcholově isomorfní, pak i $H[e/K]$ a $H'[e/K']$ jsou vrcholově isomorfní.



Obrázek 2: Příklad náhrady hyperhrany: (a) graf H, (b) graf K, (c) výsledek substituce $H[e/K]$

Na obr. 2 je příklad náhrady hyperhrany e (je to hrana ohodnocená symbolem B) v grafu H grafem K.

Dá se ukázat, že náhrada hyperhran je konfluentní a asociativní operace.

Konfluence (confluence) znamená, že pro libovolné disjunktní hypergrafy H, K_1 a K_2 a libovolné dvě různé hrany $e_1, e_2 \in E_H$ takové, že $\text{type}(e_1) = \text{type}(K_1)$ a $\text{type}(e_2) = \text{type}(K_2)$, platí, že $H[e_1/K_1][e_2/K_2] = H[e_2/K_2][e_1/K_1]$, tj. nezáleží na pořadí substitucí.

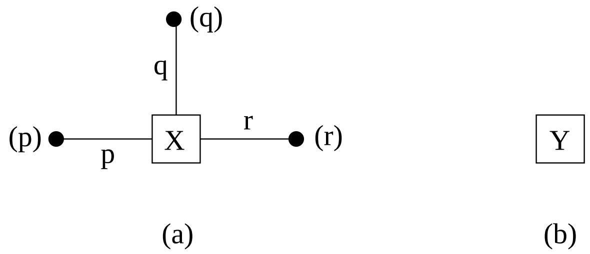
Asociativita (associativity) znamená, že pro libovolné disjunktní hypergrafy H, K_1 a K_2 a libovolné hrany $e_1 \in E_H$ a $e_2 \in E_{K_1}$ takové, že $\text{type}(e_1) = \text{type}(K_1)$ a $\text{type}(e_2) = \text{type}(K_2)$, platí, že $H[e_1/K_1][e_2/K_2] = H[e_1/K_1[e_2/K_2]]$, tj. nezáleží na tom, zda je hrana e_2 v grafu K_1 nahrazena grafem K_2 předtím nebo až potom, co je v grafu H hrana e_1 nahrazena grafem K_1.

2.4 Definice HR gramatik

Každý neterminál v HR gramatikách má přiřazen typ. ***Typ neterminálu*** je množina selektorů (podobně jako typ hrany nebo typ grafu). Existuje tedy zobrazení $\text{type} : N \to 2^{\Sigma}$, kde N je množina neterminálů, Σ je množina selektorů a kde zápis 2^{Σ} označuje množinu všech podmnožin množiny Σ. Pokud je hrana ohodnocena neterminálem, musí být hrana i neterminál stejného typu.

Zápisem $sn(X, e)$ označujeme hypergraf, tvořený jedinou hyperhranou e, která je ohodnocena neterminálem X a je stejného typu jako X. Tento hypergraf obsahuje pouze vrcholy incidentní s touto hranou, každý vrchol odpovídá právě jednomu selektoru). Všechny vrcholy grafu jsou externí a jsou ohodnoceny stejnými selektory jakými jsou označena tykadla spojující je s hranou e. Formální definice je následující: $sn(X, e) = (V, E, \text{lab}, \text{nod}, \text{ext})$, kde $V = \text{type}(X)$, $E = \{e\}$, $\text{lab}(e) = X$, $\text{nod}(e, s) = s$ a $\text{ext}(s) = s$ pro všechna $s \in \text{type}(X)$.

Na obr. 3 jsou dva příklady takovýchto hypergrafů: na obr. 3(a) hypergraf $sn(X, e)$, kde $\text{type}(X) = \{p, q, r\}$, a na obr. 3(b) hypergraf $sn(Y, e)$, kde $\text{type}(Y) = \emptyset$.



Obrázek 3: Příklady hypergrafů tvořených jedinou hyperhranou: (a) $sn(X, e)$, (b) $sn(Y, e)$

Definice 2.3 ***HR gramatika*** (HR grammar, hyperedge replacement grammar) je pětice $G = (N, T, \Sigma, P, S)$, kde

N — je abeceda neterminálních symbolů, přičemž každý neterminální symbol je určitého typu (typ je podmnožina množiny Σ), neterminály slouží k ohodnocení hran,

T — je abeceda terminálních symbolů sloužících k ohodnocení hran,

Σ — je množina (abeceda) selektorů,

P — je konečná množina přepisovacích (produkčních) pravidel, kde každé pravidlo je typu $X \to D$, kde $X \in N$, $D \in HGR_{N \cup T, \Sigma}$ a $\mathrm{type}(X) = \mathrm{type}(D)$, navíc pro každou hranu $e \in E_D$, která je ohodnocena neterminálním symbolem ($\mathrm{lab}_D(e) \in N$), platí, že tato hrana je stejného typu jako symbol jímž je ohodnocena, tj. $\mathrm{type}(e) = \mathrm{type}(\mathrm{lab}_D(e))$,

S — je počáteční symbol, $S \in N$.

□

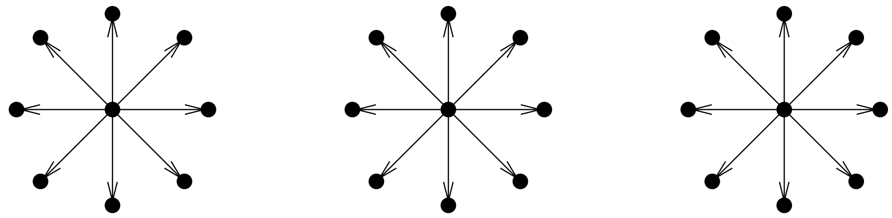
Hraně, která je ohodnocena neterminálem říkáme ***neterminální hrana***, hraně, která je ohodnocena terminálem říkáme ***terminální hrana***.

V přepisovacím pravidle $p : X \to D$ označujeme ***levou stranu*** pravidla $\mathrm{lhs}(p) = X$ (left-hand side) a ***pravou stranu*** pravidla $\mathrm{rhs}(p) = D$ (right-hand side). Dvě přepisovací pravidla $X_1 \to D_1$ a $X_2 \to D_2$ jsou ***isomorfní*** (isomorphic) pokud platí, že $X_1 = X_2$ a pokud grafy D_1 a D_2 jsou isomorfní. Přepisovací pravidlo isomorfní s některým pravidlem patřícím do P se nazývá ***kopie přepisovacího pravidla*** (production copy).

Proces přepisování podle přepisovacích pravidel gramatiky G je definován takto: Nechť $G = (N, T, \Sigma, P, S)$ je HR gramatika, nechť H a H' jsou hypergrafy patřící do $HGR_{N \cup T, \Sigma}$, nechť e je hrana grafu H a nechť $p : X \to D$ je kopie přepisovacího pravidla z gramatiky G taková, že grafy D a H jsou disjunktní. Potom pokud platí, že $\mathrm{lab}_H(e) = X$ a $H' = H[e/D]$, tj. graf H' dostaneme tak, že v grafu H nahradíme hranu e (ohodnocenou neterminálem X) grafem D, pak můžeme zapsat jeden ***derivační krok*** (derivation step) výrazem $H \Rightarrow_{e,p} H'$ nebo stručněji $H \Rightarrow H'$. Posloupnost takovýchto derivačních kroků se nazývá ***derivace*** (derivation). Derivace

$$H_0 \Rightarrow_{e_1,p_1} H_1 \Rightarrow_{e_2,p_2} \cdots \Rightarrow_{e_n,p_n} H_n.$$

se nazývá ***kreativní derivace*** (creative derivation), jestliže grafy H_0 a $\mathrm{rhs}(p_i), 1 \leq i \leq n$ jsou navzájem disjunktní. V dalším textu se omezujeme pouze na kreativní derivace. Pokud existuje kreativní derivace taková, že ve výše uvedeném zápise platí, že $H_0 = H$ a $H_n = H'$, používáme zápis $H \Rightarrow^* H'$.



Obrázek 4: Příklad grafu generovaného gramatikou G_1

Větná forma (sentential form) gramatiky G je hypergraf H takový, že $sn(S, e) \Rightarrow^* H$ pro nějaké e (tj. hypergraf H můžeme odvodit pomocí přepisovacích pravidel gramatiky G z grafu tvořeného jedinou hranou ohodnocenou počátečním neterminálním symbolem S).

Pro každý neterminál X ($X \in N$) je ***hypergrafový jazyk*** (hypergraph language) generovaný tímto neterminálem v gramatice G definován jako množina

$$L(G, X) = \{[H] : H \in HGR_{T,\Sigma} \text{ a } sn(X, e) \Rightarrow^* H \text{ pro nějaké } e\},$$

tj. jako množina všech abstraktních hypergrafů obsahujících pouze terminální hrany, jejichž konkrétní instance je možno odvodit z grafu tvořeného jedinou hranou ohodnocenou neterminálem X.

Hypergrafový jazyk generovaný gramatikou G je množina $L(G) = L(G, S)$, tj. jazyk generovaný počátečním neterminálem S. Pozn.: $L(G) \subseteq [HRG_{T,\Sigma}]$.

Třídu všech jazyků generovaných HR gramatikami označujeme HR. Gramatiky G a G' jsou ekvivalentní pokud generují tentýž jazyk, tj. $L(G) = L(G')$.

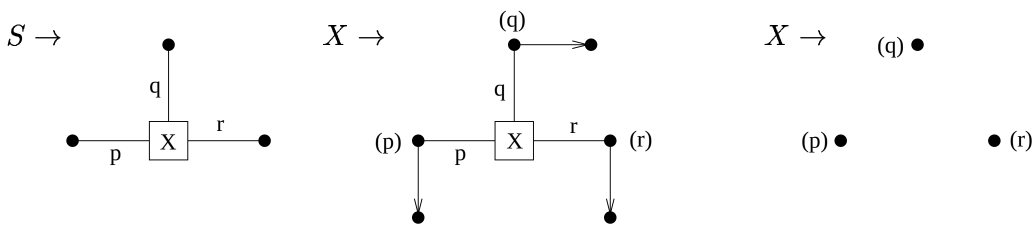
2.5 Příklady HR gramatik

Příklad 1:

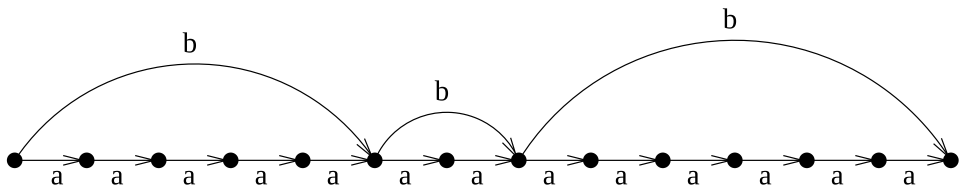
Prvním příkladem je gramatika generující grafový jazyk tvořený všemi „trojicemi hvězd“. Příklad jedné takové trojice hvězd je na obr. 4 (symboly $*$, kterými jsou ohodnoceny hrany, nejsou vyznačeny). Každá taková trojice je tvořena třemi disjunktními kopiemi jedné hvězdy. Tento grafový jazyk je generován HR gramatikou $G_1 = (N, T, \Sigma, P, S)$, kde $N = \{S, X\}$, $T = \{*\}$, $\Sigma = \{\text{sou}, \text{tar}, p, q, r\}$, $\text{type}(S) = \emptyset$, $\text{type}(X) = \{p, q, r\}$ a P obsahuje tři přepisovací pravidla nakreslená na obr. 5. Trojice hvězd, kde každá hvězda obsahuje n hran, je generována tak, že nejprve je použito první přepisovací pravidlo, poté je n-krát použito druhé přepisovací pravidlo, přičemž každé jeho použití přidá jednu hranu ke každé hvězdě, a nakonec je použito třetí přepisovací pravidlo, které odstraní neterminální hranu ohodnocenou neterminálem X.

Příklad 2:

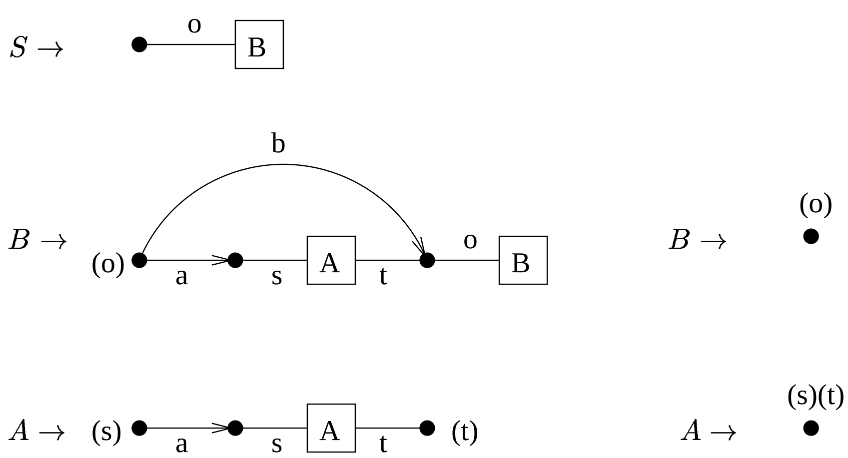
Druhým příkladem je gramatika generující všechny grafy podobného typu jako je graf na obr. 6. Je to gramatika $G_2 = (N, T, \Sigma, P, S)$, kde $N = \{S, A, B\}$, $T = \{a, b\}$, $\Sigma = \{\text{sou}, \text{tar}, \text{owner}\}$, $\text{type}(S) = \emptyset$, $\text{type}(A) = \{\text{sou}, \text{tar}\}$, $\text{type}(B) = \{\text{owner}\}$ a pravidla z množiny P jsou nakreslena na obr. 7 (na obrázku jsou selektory sou, tar a owner zkráceny na s, t a o). Na neterminály A a B se můžeme dívat tak, že neterminálem A jsou ohodnoceny obyčejné orientované hrany a neterminálem B jsou ohodnoceny vrcholy.



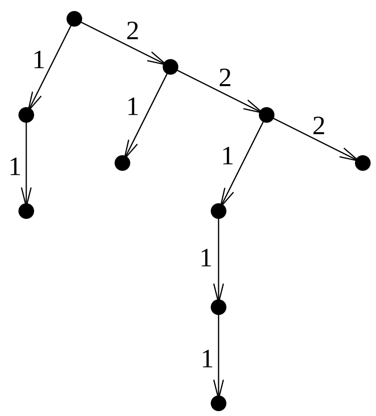
Obrázek 5: Přepisovací pravidla gramatiky G_1



Obrázek 6: Příklad grafu generovaného gramatikou G_2



Obrázek 7: Přepisovací pravidla gramatiky G_2



Obrázek 8: Příklad grafu generovaného gramatikou G_3

Příklad 3:

Třetím příkladem je gramatika G_3 generující množinu všech „hřebenů". Příklad jednoho takového hřebenu je na obr. 8, jedná se v podstatě o stromy určitého typu. Přepisovací pravidla gramatiky G_3 jsou nakreslena na obr. 9.

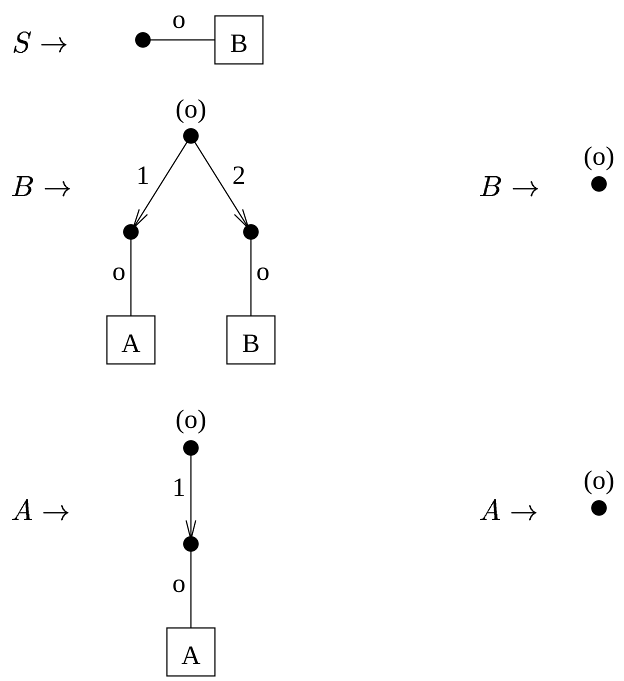
2.6 Derivační stromy pro HR gramatiky

Pro HR gramatiky platí (díky tomu, že náhrada hyperhran je asociativní a konfluentní operace) následující věta, která je uvedena bez důkazu:

Věta 2.4 Nechť $G = (N, T, \Sigma, P, S)$ je HR gramatika, H je hypergraf z množiny $HGR_{N \cup T, \Sigma}$, K je hypergraf z množiny $HGR_{T,\Sigma}$ a $e_1, \ldots, e_n$ jsou všechny neterminální hrany hypergrafu H, přičemž každá hrana e_i je ohodnocena neterminálním symbolem X_i. Potom $H \Rightarrow^* K$ právě když existují hypergrafy $K_1, \ldots, K_n \in HGR_{T,\Sigma}$ disjunktní s grafem H takové, že $K = H[e_1/K_1] \cdots [e_n/K_n]$ a $sn(X_i, e_i) \Rightarrow^* K_i$ pro všechna $1 \leq i \leq n$. Navíc se délka derivace $H \Rightarrow^* K$ rovná součtu délek derivací $sn(X_i, e_i) \Rightarrow^* K_i$. $\square$

Pojem derivační strom pro HR gramatiky je možné definovat podobně jako pro klasické bezkontextové gramatiky. Jsou zde však určité rozdíly. Jednak jednotlivé vrcholy derivačního stromu nejsou ohodnoceny symboly, ale pravidly dané gramatiky. Dále vzhledem k tomu, že neexistuje žádné jednoznačné pořadí hran v grafu (tak jako existuje jednoznačné pořadí symbolů v řetězci), je nutné neterminální hrany pravých stran pravých stran pravidel nějakým (libovolným) způsobem uspořádat.

Definice 2.5 Nechť $G = (N, T, \Sigma, P, S)$ je HR gramatika. ***Derivační strom*** (derivation tree) pro neterminál $X \in N$ je strom, jehož vrcholy jsou ohodnoceny pravidly z množiny P. Pro pravidlo p, kterým je ohodnocen kořen stromu musí platit, že $\text{lhs}(p) = X$. Potomky kořene jsou derivační stromy $t_1, \ldots, t_n$, kde t_i je derivační strom pro neterminál X_i, což je neterminál, kterým je ohodnocena i-tá neterminální hrana pravé strany pravidla p ($\text{rhs}(p)$ má právě n neterminálních hran). $\square$



Obrázek 9: Přepisovací pravidla gramatiky G_3

3 C-edNCE gramatiky

C-edNCE gramatiky neboli gramatiky s náhradou vrcholů (NR grammars, node replacement grammars) generují běžné orientované grafy (nikoli tedy hypergrafy) s ohodnocenými vrcholy a hranami. Při používání přepisovacích pravidel je vždy nahrazen jeden vrchol grafu novým grafem, přičemž oba grafy jsou spojeny přidáním nových hran podle určitých pravidel.

3.1 Grafy

V C-edNCE gramatikách se berou v úvahu pouze grafy bez smyček (tj. bez hran, které začínají i končí v tomtéž vrcholu) a bez násobných hran, což znamená, že z jednoho vrcholu do druhého nemůže vést více než jedna hrana ohodnocená stejným symbolem (ovšem z jednoho vrcholu do druhého může vést více hran z nichž každá je ohodnocena jiným symbolem).

Nechť Σ je abeceda symbolů jimiž mohou být ohodnoceny vrcholy a nechť Γ je abeceda symbolů jimiž mohou být ohodnoceny hrany. ***Graf*** (graph) nad Σ a Γ je trojice $H = (V, E, \lambda)$, kde V je konečná množina vrcholů, $E \subseteq \{(v, \gamma, w) : v, w \in V, v \neq w, \gamma \in \Gamma\}$ je konečná množina hran (hrana vede z vrcholu v, do vrcholu w a je ohodnocena symbolem γ) a $\lambda : V \to \Sigma$ je ohodnocení vrcholů. Jednotlivé složky grafu H mohou být rovněž označeny jako V_H, E_H a λ_H. Pokud je graf neorientovaný, je relace E symetrická, to znamená, že $(v, \gamma, w) \in E \iff (w, \gamma, v) \in E$. Pokud vrcholy grafu nejsou ohodnoceny, pak $\Sigma = \{\#\}$ (tj. všechny vrcholy budou ohodnoceny symbolem $\#$), pokud hrany grafu nejsou ohodnoceny, pak $\Gamma = \{*\}$ (tj. všechny hrany budou ohodnoceny symbolem $*$).

Dva grafy H a K jsou ***isomorfní*** (isomorphic) pokud existuje bijekce $f : V_H \to V_K$ taková, že $E_K = \{(f(v), \gamma, f(w)) : (v, \gamma, w) \in E_H\}$ a pro všechny vrcholy $v \in V_H$ platí $\lambda_K(f(v)) = \lambda_H(v)$. Množina všech grafů isomorfních s H se označuje $[H]$. $[H]$ je „abstraktní“

graf, zatímco H je „konkrétní“ instance tohoto abstraktního grafu.

Množina všech (konkrétních) grafů nad Σ a Γ se označuje $GR_{\Sigma,\Gamma}$ a množina všech abstraktních grafů nad Σ a Γ se označuje $[GR_{\Sigma,\Gamma}]$. Podmnožina množiny $[GR_{\Sigma,\Gamma}]$ se nazývá ***grafový jazyk*** (graph language).

Přepisovací pravidla C-edNCE gramatik jsou typu $X \to (D, C)$, kde X je neterminální symbol (je to symbol, kterým mohou být ohodnoceny vrcholy grafu), D je graf a C je množina instrukcí pro připojení grafu D k původnímu grafu. Použití takového pravidla při náhradě vrcholu v (ohodnoceného symbolem X) v grafu H spočívá v odstranění vrcholu v z grafu H, přidání grafu D ke grafu H (D a H musí být disjunktní) a přidání nových hran mezi grafy D a H podle připojovacích instrukcí C. Na dvojici (D, C) lze pohlížet jako na jediný objekt, v podstatě se jedná o graf s určitým rozhraním, pomocí kterého může být připojen k jinému grafu. Následuje formální definice tohoto objektu.

Definice 3.1 Nechť Σ je abeceda symbolů pro ohodnocení vrcholů a nechť Γ je abeceda symbolů pro ohodnocení hran. Potom ***graf s připojením*** (graph with embedding) nad Σ a Γ je čtveřice $K = (V, E, \lambda, C)$, kde $(V, E, \lambda) \in GR_{\Sigma,\Gamma}$ a $C \subseteq \Sigma \times \Gamma \times \Gamma \times V \times \{\text{in}, \text{out}\}$. C je ***relace připojení*** (connection relation) grafu K. Prvek $(\sigma, \beta, \gamma, x, d)$ patřící do množiny C (přičemž $\sigma \in \Sigma$, $\beta, \gamma \in \Gamma$, $x \in V$ a $d \in \{\text{in}, \text{out}\}$) se nazývá ***připojovací instrukce*** (connection instruction) grafu K. □

Pro jednotlivé komponenty grafu K je možné použít označení V_K, E_K, λ_K a C_K. Kvůli lepší čitelnosti se pro připojovací instrukci $(\sigma, \beta, \gamma, x, d)$ používá zápis $(\sigma, \beta/\gamma, x, d)$.

Pozn.: V následujícím textu se kvůli stručnosti někdy používá pojem graf místo pojmu graf s připojením.

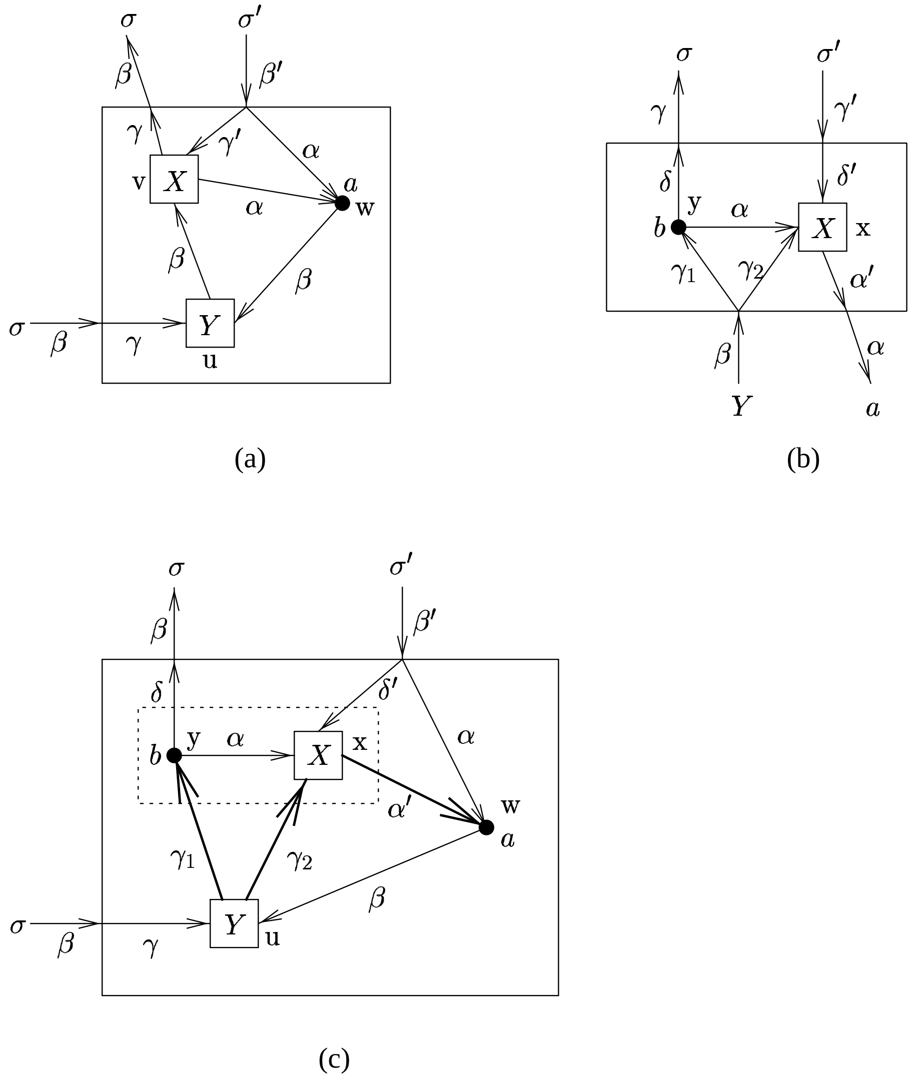
Intuitivní význam připojovací instrukce $(\sigma, \beta/\gamma, x, \text{out}) \in C_K$ je následující: Pokud nahrazujeme v grafu H vrchol v grafem K a v grafu H existovala hrana ohodnocená symbolem β vedoucí z vrcholu v do vrcholu w, který je ohodnocen symbolem σ, pak po nahrazení vrcholu v grafem K přidáme k výslednému grafu hranu z vrcholu x ($x \in V_K$) do vrcholu w ($w \in V_H$) ohodnocenou symbolem γ. Podobně pokud máme připojovací instrukci $(\sigma, \beta/\gamma, x, \text{in}) \in C_K$ a v grafu H existovala hrana z vrcholu w do vrcholu v ohodnocená symbolem β, přidáme do výsledného grafu hranu z vrcholu w do vrcholu x ohodnocenou symbolem γ.

Grafy s připojením H a K jsou ***isomorfní*** (isomorphic) pokud existuje bijekce f taková, že grafy (V_H, E_H, λ_H) a (V_K, E_K, λ_K) jsou isomorfní podle výše uvedené definice a pokud platí, že $C_K = \{(\sigma, \beta/\gamma, f(x), d) : (\sigma, \beta/\gamma, x, d) \in C_H\}$. Grafy H a K jsou ***disjunktní*** (disjoint) pokud nemají žádný společný vrchol, tj. pokud $V_H \cap V_K = \emptyset$.

Množinu všech konkrétních grafů s připojením nad Σ a Γ označujeme $GRE_{\Sigma,\Gamma}$ a množinu všech abstraktních grafů s připojením nad Σ a Γ označujeme $[GRE_{\Sigma,\Gamma}]$. Na „obyčejné“ grafy se můžeme dívat jako na grafy s připojením, které neobsahují žádné připojovací instrukce (tj. $C = \emptyset$). Proto platí vztah $GR_{\Sigma,\Gamma} \subseteq GRE_{\Sigma,\Gamma}$.

Na obr. 10(a) je nakreslen graf s připojením H patřící do $GRE_{\Sigma,\Gamma}$ kde

$$
\begin{aligned}
&\Sigma \;\; = \{X, Y, a, b, \sigma, \sigma'\},\\
&\Gamma \;\; = \{\alpha, \alpha', \beta, \beta', \gamma, \gamma', \gamma_1, \gamma_2, \delta, \delta'\},\\
&V_H = \{u, v, w\},\\
&\lambda_H(u) = Y,\ \lambda_H(v) = X,\ \lambda_H(w) = a,\\
&E_H = \{(u, \beta, v), (v, \alpha, w), (w, \beta, u)\},\\
&C_H = \{(\sigma, \beta/\gamma, u, \text{in}), (\sigma, \beta/\gamma, v, \text{out}), (\sigma', \beta'/\gamma', v, \text{in}), (\sigma', \beta'/\alpha, w, \text{in})\}.
\end{aligned}
$$



Obrázek 10: Dva grafy s připojením a výsledek substituce: (a) graf H, (b) graf K, (c) graf $H[v/K]$

Na obr. 10(b) je nakreslen graf s připojením K rovněž patřící do $GRE_{\Sigma,\Gamma}$ kde Σ a Γ jsou stejné jako u grafu H a

$V_K = \{x, y\}$,
$\lambda_K(x) = X$, $\lambda_K(y) = b$,
$E_K = \{(y, \alpha, x)\}$,
$C_K = \{(\sigma, \gamma/\delta, y, \mathrm{out}), (\sigma', \gamma'/\delta', x, \mathrm{in}), (Y, \beta/\gamma_1, y, \mathrm{in}), (Y, \beta/\gamma_2, y, \mathrm{in}), (a, \alpha/\alpha', x, \mathrm{out})\}$

Z obrázků je patrné jakým způsobem je možné grafy s připojením graficky znázornit. Grafy jsou znázorněny obvyklým způsobem s tím rozdílem, že jsou nakresleny uvnitř většího obdélníka a že vrcholy, které jsou ohodnoceny neterminálním symbolem, jsou zobrazeny jako čtverce. Připojovací instrukce jsou znázorněny jako šipky vedoucí z nebo do některého vrcholu přičemž jejich druhý konec se vždy nachází mimo obdélník, který ohraničuje graf, a je ohodnocen některým ze symbolů pro ohodnocení vrcholů. Každá taková šipka je ohodnocena dvěma symboly: jedním vně obdélníka (je to „starý“ symbol) a druhým uvnitř obdélníka (je to „nový“ symbol). Šipky mimo obdélník se stejnými symboly je možné spojit v jedinou šipku. Pokud graf neobsahuje žádné připojovací instrukce, není nutné vyznačovat ohraničující obdélník.

3.2 Náhrada vrcholů

Následuje formální definice náhrady vrcholu grafem s připojením.

Definice 3.2 Nechť $H, K \in GRE_{\Sigma,\Gamma}$ jsou dva disjunktní grafy s připojením a nechť v je vrchol grafu H. ***Náhradou*** neboli substitucí (substitution) vrcholu v v grafu H grafem K dostaneme graf s připojením (V, E, λ, C) patřící do $GRE_{\Sigma,\Gamma}$, kde

$$
\begin{aligned}
V &= (V_H - \{v\}) \cup V_K, \\
E &= \{(x,\gamma,y) \in E_H : x \neq v, y \neq v\} \cup E_K \\
&\quad \cup \{(w,\gamma,x) : \exists \beta \in \Gamma : (w,\beta,v) \in E_H, (\lambda_H(w), \beta/\gamma, x, \mathrm{in}) \in C_K\} \\
&\quad \cup \{(x,\gamma,w) : \exists \beta \in \Gamma : (v,\beta,w) \in E_H, (\lambda_H(w), \beta/\gamma, x, \mathrm{out}) \in C_K\}, \\
\lambda(x) &= \begin{cases} \lambda_H(x) \text{ pokud } x \in V_H - \{v\} \\ \lambda_K(x) \text{ pokud } x \in V_K \end{cases}, \\
C &= \{(\sigma, \beta/\gamma, x, d) \in C_H : x \neq v\} \\
&\quad \cup \{(\sigma, \beta/\gamma, x, d) : \exists \gamma \in \Gamma((\sigma, \beta/\gamma, v, d) \in C_H \wedge (\sigma, \gamma/\delta, x, d) \in C_K)\}.
\end{aligned}
$$

Tento výsledný graf s připojením označujeme zápisem $H[v/K]$. □

Příklad takové náhrady vrcholu je na obr. 10(c), kde je nakreslen výsledek substituce $H[v/K]$, kde H a K jsou grafy na obr. 10(a) a obr. 10(b). Ve výsledném grafu jsou pro názornost ohraničeny vrcholy vzniklé přidáním grafu K přerušovanou čárou a nově přidané hrany jsou vyznačeny tučně. Výsledkem je graf (V, E, λ, C), kde

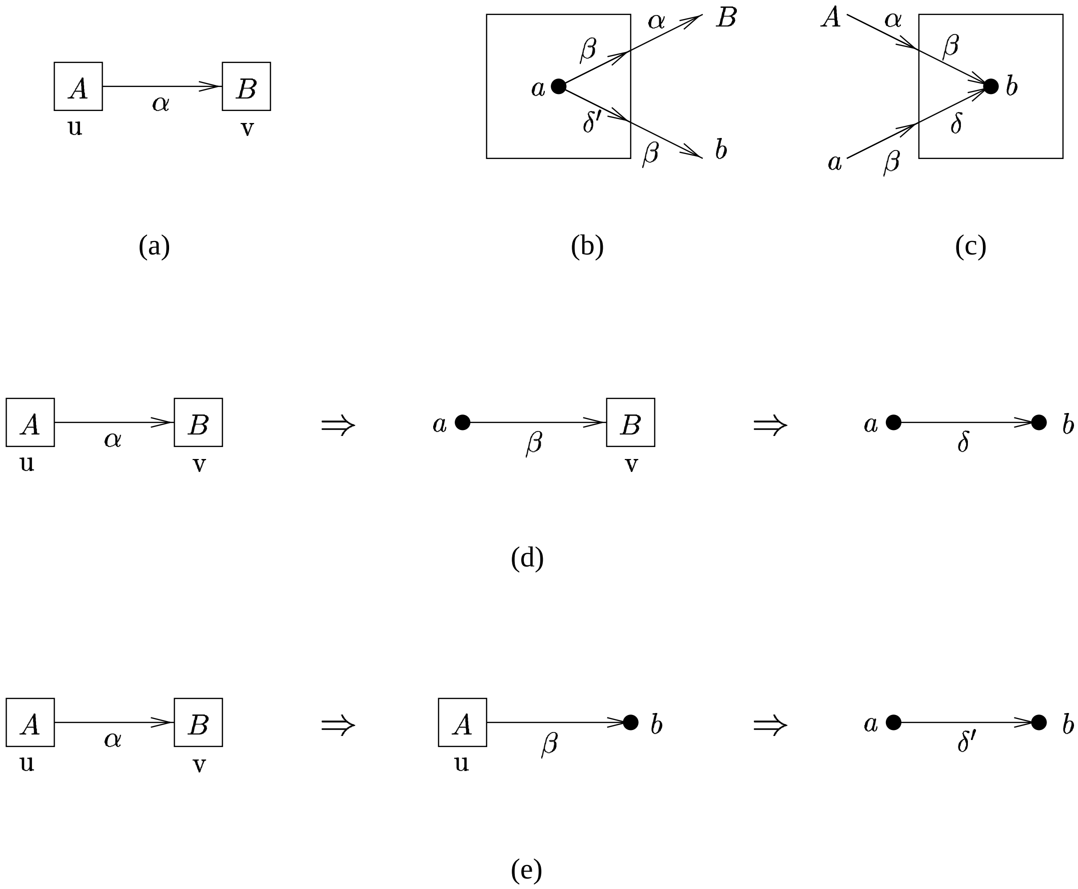
$$
\begin{aligned}
&V = \{u, w, x, y\}, \\
&\lambda(u) = Y,\ \lambda(w) = a,\ \lambda(x) = X,\ \lambda(y) = b, \\
&E = \{(w,\beta,u), (y,\alpha,x), (x,\alpha',w), (u,\gamma_1,y), (u,\gamma_2,x)\}, \\
&C = \{(\sigma', \beta'/\alpha, w, \mathrm{in}), (\sigma, \beta/\gamma, u, \mathrm{in}), (\sigma, \beta/\delta, y, \mathrm{out}), (\sigma', \beta'/\delta', x, \mathrm{in})\}.
\end{aligned}
$$

Je dobré si všimnout, že rozhraní grafu H mimo ohraničující obdélník zůstalo po substituci nezměněno.

Aby se operace náhrady vrcholu grafem mohla stát základem bezkontextové grafové gramatiky, je nutné, aby tato operace byla ***asociativní*** (associative) a ***konfluentní*** (confluent). Dá se ukázat, že tato operace je asociativní, což znamená, že pokud H, K_1 a K_2 jsou disjunktní grafy s připojením, v_1 je vrchol grafu H a v_2 je vrchol grafu K_1, pak platí, že $H[v_1/K_1][v_2/K_2] = H[v_1/K_1[v_2/K_2]]$, tj. nezáleží na tom jestli je nejprve nahrazen vrchol v_1 v grafu H grafem K_1 a pak vrchol v_2 v grafu K_1 grafem K_2 nebo naopak.

Jak ukazuje následující příklad, nahrazování vrcholů ***není konfluentní***. Když vezmeme například grafy H, K a M, jaké jsou nakresleny na obr. 11, je z obrázku patrné, že výsledek substituce závisí na pořadí nahrazování jednotlivých vrcholů, že tedy $H[u/K][v/M]$ a $H[v/M][u/K]$ jsou dva různé grafy lišící se ohodnocením hrany.

Jak vyplývá z tohoto příkladu, není nahrazování vrcholů v obecném případě konfluentní, proto by grafová gramatika obsahující přepisovací pravidla založená na této operaci nemusela



Obrázek 11: Příklad ukazující, že náhrada vrcholů není konfluentní operace: (a) graf H, (b) graf K, (c) graf M, (d) derivace $H \Rightarrow H[u/K] \Rightarrow H[u/K][v/M]$, (e) derivace $H \Rightarrow H[v/M] \Rightarrow H[v/M][u/K]$

být v obecném případě bezkontextová. Proto je nutné explicitně zajistit, aby operace nahrazování vrcholů byla pro množinu pravidel dané gramatiky konfluentní. K tomu složí následující definice.

Definice 3.3 Nechť P je množina přepisovacích (produkčních) pravidel typu $X \to D$, kde $X \in \Sigma$ a $D \in GRE_{\Sigma,\Gamma}$. Říkáme, že množina pravidel P je ***konfluentní*** (confluent), jestliže pro libovolné disjunktní grafy $H, D_1, D_2 \in GRE_{\Sigma,\Gamma}$ a libovolné dva vrcholy $v_1, v_2 \in V_H$, kde $v_1 \neq v_2$, platí, že pokud v P jsou pravidla $\lambda_H(v_1) \to D'_1$ a $\lambda_H(v_2) \to D'_2$, kde D'_1 a D'_2 jsou grafy isomorfní s grafy D_1 a D_2, pak $H[v_1/D_1][v_2/D_2] = H[v_2/D_2][v_1/D_1]$. □

Jestli je daná množina přepisovacích pravidel konfluentní se dá snadno ověřit pomocí následujícího tvrzení, které je uvedeno bez důkazu:

Tvrzení 3.4 Nechť P je množina přepisovacích pravidel typu $X \to D$, kde $X \in \Sigma$ a $D \in GRE_{\Sigma,\Gamma}$. Pak je P konfluentní právě když pro libovolná dvě pravidla $X_1 \to D_1$ a $X_2 \to D_2$ patřící do P, pro libovolné dva vrcholy $x_1 \in V_{D_1}$ a $x_2 \in V_{D_2}$ a pro libovolné dva symboly $\alpha, \delta \in \Gamma$ platí následující ekvivalence:

$$\exists \beta \in \Gamma : (X_2, \alpha/\beta, x_1, \mathrm{out}) \in C_{D_1} \;\wedge\; (\lambda_{D_1}(x_1), \beta/\delta, x_2, \mathrm{in}) \in C_{D_2}$$
$$\Longleftrightarrow$$
$$\exists \gamma \in \Gamma : (X_1, \alpha/\gamma, x_2, \mathrm{in}) \in C_{D_2} \;\wedge\; (\lambda_{D_2}(x_2), \gamma/\delta, x_1, \mathrm{out}) \in C_{D_1}.$$

□

3.3 Definice C-edNCE gramatik

Význam názvu C-edNCE gramatik je následující: 'C' znamená konfluentní (confluent), 'NCE' znamená, že připojování závisí na sousedech vrcholu (neighbourhood controlled embedding), 'd' znamená, že gramatika generuje orientované grafy (directed graphs) a 'e' znamená, že ohodnocení hran se může při přepisování měnit (dynamic edge relabeling).

Definice 3.5 ***C-edNCE gramatika*** (C-edNCE grammar) nebo též ***NR gramatika*** (NR grammar, node replacement grammar) je šestice $G = (\Sigma, \Delta, \Gamma, \Omega, P, S)$, kde

Σ — je abeceda symbolů pro označení vrcholů,
Δ — je abeceda terminálních symbolů pro označení vrcholů, $\Delta \subseteq \Sigma$,
Γ — je abeceda symbolů pro označení hran,
Ω — je abeceda terminálních symbolů pro označení hran, $\Omega \subseteq \Gamma$,
P — je konečná množina přepisovacích (produkčních) pravidel typu $X \to D$, kde $X \in \Sigma - \Delta$ a $D \in GRE_{\Sigma,\Gamma}$,
S — je počáteční neterminální symbol, $S \in \Sigma - \Delta$.

Navíc musí mít gramatika následující tři vlastnosti:

(1) Množina přepisovacích pravidel P musí být konfluentní.

(2) Pro každé přepisovací pravidlo $X \to D \in P$ platí, že pokud $(x, \gamma, y) \in E_D$ přičemž $\lambda_D(x), \lambda_D(y) \in \Delta$, pak $\gamma \in \Omega$; dále platí, že pokud $(\sigma, \beta/\gamma, x, d) \in C_D$, $\sigma \in \Delta$ a $\lambda_D(x) \in \Delta$, pak $\gamma \in \Omega$.

(3) Pokud $S \to D \in P$, pak $C_D = \emptyset$.

$\square$

Vlastnost (1) zajišťuje, že přepisovací pravidla gramatiky jsou konfluentní, což je důležité proto, aby se jednalo skutečně o bezkontextovou gramatiku. Vlastnost (2) zajišťuje, že libovolný graf vygenerovaný gramatikou jehož všechny vrcholy jsou ohodnoceny terminálními symboly z abecedy Δ, bude mít všechny hrany ohodnoceny terminálními symboly z abecedy Ω. Vlastnost (3) zajišťuje, že vygenerovaný graf neobsahuje žádné připojovací instrukce.

Symboly patřící do množiny $\Sigma - \Delta$ se nazývají neterminální symboly pro označení vrcholů a symboly patřící do množiny $\Gamma - \Omega$ se nazývají neterminální symboly pro označení hran. Vrcholům, které jsou ohodnoceny terminálními nebo neterminálními symboly, říkáme terminální nebo neterminální vrcholy a hranám, které jsou ohodnoceny terminálními nebo neterminálními symboly, říkáme terminální nebo neterminální hrany. Levou stranu přepisovacího pravidla $p : X \to D$ označujeme $\mathrm{lhs}(p) = X$ (left-hand side), pravou stranu označujeme $\mathrm{rhs}(p) = D$ (right-hand side). Dvě přepisovací pravidla $X_1 \to D_1$ a $X_2 \to D_2$ jsou ***isomorfní*** (isomorphic) pokud platí, že $X_1 = X_2$ a pokud grafy D_1 a D_2 jsou isomorfní. Přepisovací pravidlo isomorfní s některým přepisovacím pravidlem z množiny P se nazývá ***kopie přepisovacího pravidla*** (production copy).

Proces přepisování v dané C-edNCE gramatice je definován následujícím způsobem: Nechť $G = (\Sigma, \Delta, \Gamma, \Omega, P, S)$ je C-edNCE gramatika, H a H' jsou grafy patřící do $GRE_{\Sigma,\Gamma}$, $v \in V_H$ je vrchol grafu H a $p : X \to D$ je kopie přepisovacího pravidla z gramatiky G taková, že grafy D a H jsou disjunktní. Pak zapisujeme jeden ***derivační krok*** (derivation step) výrazem $H \Rightarrow_{v,p} H'$ nebo stručněji $H \Rightarrow H'$, jestliže je vrchol v je ohodnocen neterminálem X, tj. $\lambda_H(v) = X$, a jestliže po nahrazení vrcholu v v grafu H grafem D dostaneme graf H', tj. $H' = H[v/D]$. Posloupnost takovýchto derivačních kroků nazýváme ***derivace*** (derivation). Derivace

$$H_0 \Rightarrow_{v_1,p_1} H_1 \Rightarrow_{v_2,p_2} \cdots \Rightarrow_{v_n,p_n} H_n,$$

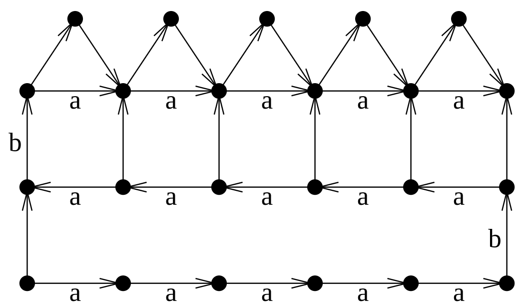
se nazývá ***kreativní derivace*** (creative derivation), jestliže grafy H_0 a $\mathrm{rhs}(p_i)$, kde $1 \leq i \leq n$, jsou navzájem disjunktní. V dalším popisu se omezujeme pouze na kreativní derivace. Pokud je možné dostat nějakou derivací z grafu H graf H' (tj. ve výše uvedeném zápise je $H_0 = H$ a $H_n = H'$), zapisujeme tuto skutečnost výrazem $H \Rightarrow^* H'$. Pokud platí, že $H \Rightarrow^* H'$ a navíc $H \neq H'$, pak používáme zápis $H \Rightarrow^+ H'$.

Zápisem $sn(X, x)$ označujeme graf tvořený jediným vrcholem x ohodnoceným symbolem X, který neobsahuje žádné hrany a má všechny možné připojovací instrukce, ve kterých je starý a nový symbol pro ohodnocení hrany totožný. Formálně zapsáno $sn(X, x) = (V, E, \lambda, C)$, kde $V = \{x\}$, $E = \emptyset$, $\lambda(x) = X$ a $C = \{(\sigma, \gamma/\gamma, x, d) : \sigma \in \Sigma, \gamma \in \Gamma, d \in \{\mathrm{in}, \mathrm{out}\}\}$.

Větná forma (sentential form) gramatiky G je graf H takový, že $sn(S, x) \Rightarrow^* H$ pro nějaké x. Díky vlastnosti (3) v definici 3.5 platí, že pokud $sn(S, x) \Rightarrow^+ H$, pak je H obyčejný graf (bez připojovacích instrukcí), tj. $H \in GR_{\Sigma,\Gamma}$. Jazyk generovaný v gramatice G neterminálním symbolem $X \in \Sigma - \Delta$ je množina

$$L(G, X) = \{[H] : H \in GRE_{\Delta,\Omega} \text{ a } sn(X, x) \Rightarrow^* H \text{ pro nějaké } x\},$$

tj. množina všech abstraktních grafů, které jsou tvořeny pouze terminálními vrcholy a terminálními hranami a které je možné získat derivací z grafu tvořeného jediným vrcholem ohodnoceným symbolem X. ***Grafový jazyk*** (graph language) generovaný gramatikou G je jazyk generovaný počátečním symbolem, tj. $L(G) = L(G, S) \subseteq [GR_{\Delta,\Omega}]$. Třídu grafových jazyků generovaných C-edNCE gramatikami označujeme C-edNCE.



Obrázek 12: Příklad grafu generovaného gramatikou G_1

3.4 Příklady C-edNCE gramatik

Příklad 1:

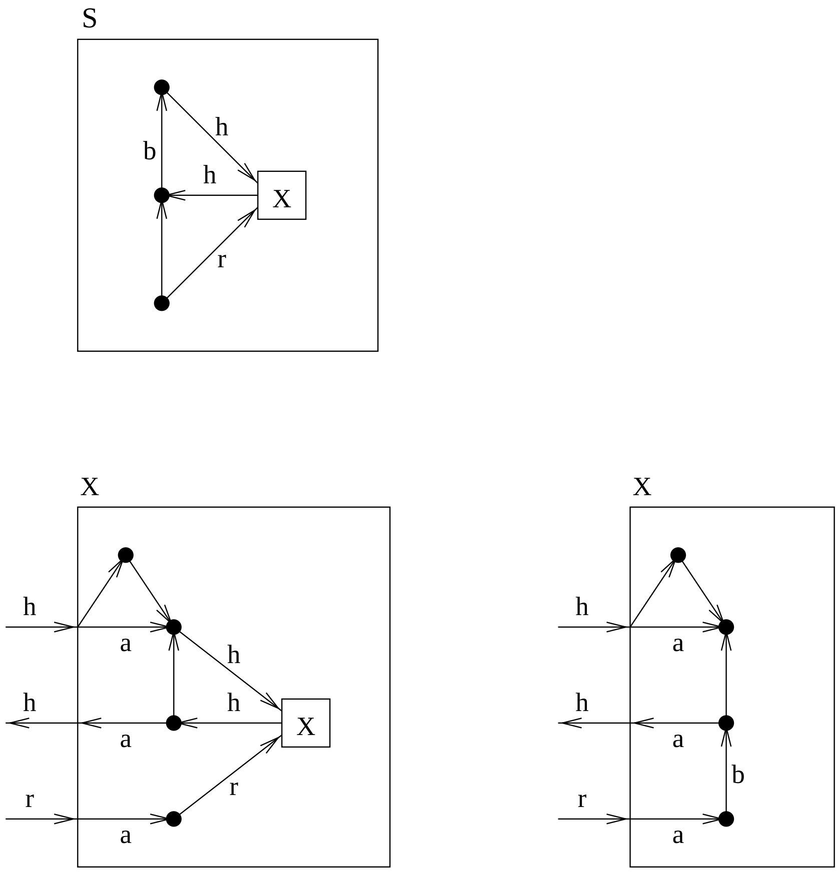
Prvním příkladem je gramatika G_1 generující množinu všech „ulic“. Příklad jedné takové ulice je na obr. 12. Hrany, které na obrázku nejsou ohodnoceny, jsou ohodnoceny symbolem $*$, vrcholy, které na obrázku nejsou ohodnoceny, jsou ohodnoceny symbolem $\#$. Gramatika G_1 je definována následujícím způsobem: $G_1 = (\Sigma, \Delta, \Gamma, \Omega, P, S)$, kde $\Sigma = \{S, X, \#\}$, $\Delta = \{\#\}$, $\Gamma = \{h, r, a, b, *\}$, $\Omega = \{a, b, *\}$ a P obsahuje pravidla nakreslená na obr. 13. Levá strana pravidel je uvedena vždy vlevo nahoře nad obdélníkem, který ohraničuje pravou stranu pravidla. Gramatika generuje ulici tvořenou n domy, kde $n \geq 1$. Nejprve je vždy použito první pravidlo, poté $n-1$ krát druhé pravidlo, přičemž každým jeho použitím se přidá jeden dům, a nakonec je přidán pomocí třetího pravidla poslední dům.

Příklad 2:

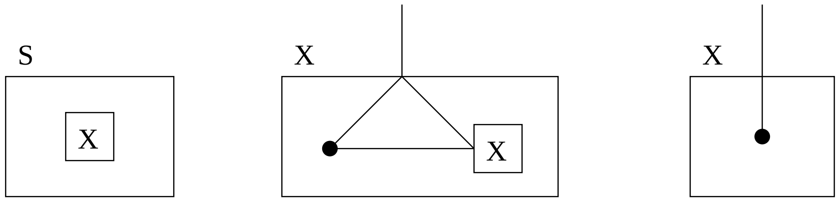
Druhým příkladem je gramatika G_2, která generuje množinu všech úplných neorientovaných grafů K_n, kde $n \geq 1$, tj. grafů ve kterých je spojen hranou každý vrchol s každým. Pravidla gramatiky jsou zobrazena na obr. 14. Jedna hrana v neorientovaném grafu odpovídá vždy dvěma hranám v orientovaném grafu a je zobrazena jako neorientovaná čára a podobně i dvě připojovací instrukce $(\sigma, \beta/\gamma, x, \mathrm{in})$ a $(\sigma, \beta/\gamma, x, \mathrm{out})$ jsou zobrazeny jako jedna neorientovaná čára. Stejně jako v předchozím případě jsou na obrázku neohodnocené hrany ohodnoceny symbolem $*$ a neohodnocené vrcholy ohodnoceny symbolem $\#$. Větná forma vygenerovaná touto gramatikou je tvořena úplným grafem, v němž je nejvýše jeden vrchol ohodnocen neterminálem X.

Příklad 3:

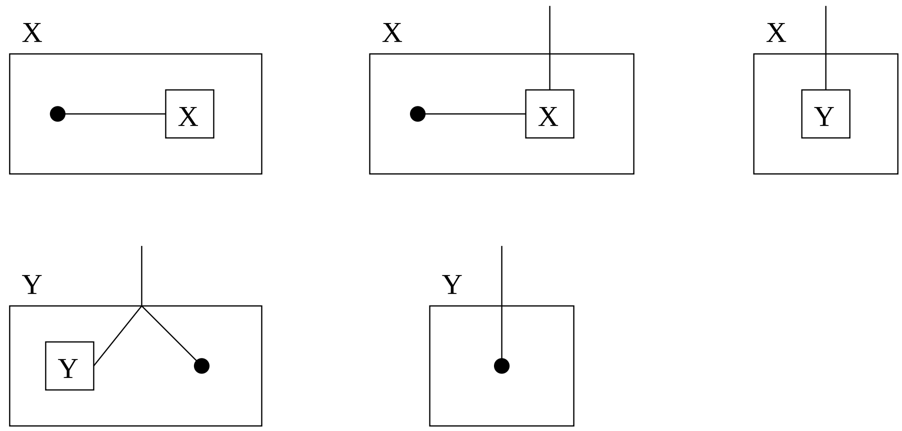
Třetím příkladem je gramatika G_3 generující množinu všech úplných neorientovaných bipartitních grafů $K_{m,n}$, kde $m, n \geq 1$, což jsou grafy, kde vrcholy jsou rozděleny do dvou skupin, přičemž každý vrchol z jedné skupiny je vždy spojen hranou s každým vrcholem z druhé skupiny, přičemž žádný vrchol není spojen s vrcholem ze stejné skupiny. Pravidla gramatiky jsou nakreslena na obr. 15, přičemž byly použity stejné konvence jako v předchozím příkladě.



Obrázek 13: Gramatika G_1 generující množinu všech „ulic“



Obrázek 14: Gramatika G_2 generující množinu všech úplných grafů



Obrázek 15: Gramatika G_3 generující množinu všech úplných bipartitních grafů

3.5 Derivační stromy pro C-edNCE gramatiky

Díky tomu, že v případě C-edNCE gramatik je náhrada vrcholů asociativní a konfluentní operace, platí následující věta, která je obdobou věty 2.4, která platí pro HR gramatiky (věta je uvedena bez důkazu):

Věta 3.6 Nechť $G = (\Sigma, \Delta, \Gamma, \Omega, P, S)$ je C-edNCE gramatika, H je graf patřící do množiny $GRE_{\Sigma,\Gamma}$, K je graf patřící do množiny $GRE_{\Delta,\Omega}$ a $v_1, \ldots, v_n$ jsou všechny neterminální vrcholy grafu H, přičemž každý vrchol v_i je ohodnocen neterminálním symbolem X_i. Potom $H \Rightarrow^* K$ právě když existují grafy $K_1, \ldots, K_n$ patřící do množiny $GRE_{\Delta,\Omega}$ a disjunktní s grafem H takové, že $K = H[v_1/K_1] \cdots [v_n/K_n]$ a $sn(X_i, v_i) \Rightarrow^* K_i$ pro všechna $1 \leq i \leq n$. Navíc platí, že délka derivace $H \Rightarrow^* K$ se rovná součtu délek derivací $sn(X_i, v_i) \Rightarrow^* K_i$. □

Pojem derivační strom je definován analogicky jako u HR gramatik. Opět je třeba nějakým způsobem (libovolně) uspořádat neterminální vrcholy pravých stran přepisovacích pravidel.

Definice 3.7 Nechť $G = (\Sigma, \Delta, \Gamma, \Omega, P, S)$ je C-edNCE gramatika. ***Derivační strom*** (derivation tree) pro neterminál $X \in \Sigma - \Delta$ je strom, jehož vrcholy jsou ohodnoceny pravidly z množiny P. Pro pravidlo p, kterým je ohodnocen kořen stromu musí platit $\text{lhs}(p) = X$. Potomky kořene jsou derivační stromy $t_1, \ldots, t_n$, kde t_i je derivační strom pro X_i, což je neterminál, kterým je ohodnocen i-tý neterminální vrchol $\text{rhs}(p)$ ($\text{rhs}(p)$ má právě n neterminálních vrcholů). □

Obsah