Slova a volná pologrupa

Nechť Σ je nějaká množina symbolů, kterou budeme nazývat *abeceda* a její prvky *písmena*. Posloupnost symbolů z abecedy nazýváme slovo a namísto $(a_1, a_2, \dots, a_n)$ píšeme prostě $a_1a_2 \cdots a_n$. Je dobré si všimnout, že tímto zápisem ztotožňujeme písmena, např. a, a slova délky jedna, např. (a). Množinu všech neprázdných slov označujeme Σ^+, přidáním prázdného slova dostáváme množinu označovanou Σ^*. Na slovech je přirozeně definována operace *zřetězení*, která např. ze slov ab, ba vytvoří slovo $abba$. Zřetězení je natolik samozřejmou operací, že může zakrývat některá základní fakta. Pokud si operaci zřetězení neoznačíme žádným symbolem, není např. možné srozumitelně a krátce zapsat vztah mezi slovy ab, ba a $abba$ vyjádřený v předchozí větě. Pokud to bude vhodné, budeme proto zřetězení značit znakem násobení. Pak můžeme napsat $ab \cdot ba = abba$. Znak násobení se podobně jako v aritmetice často vynechává. V našem případe stačí znak vynechat, abychom dostali výsledek. Slovo $abba$ lze chápat jako výsledek zřetězení jiných slov několika různými způsoby, říkáme, že ho lze různým způsobem *faktorizovat*. Nejzákladnější je faktorizace do písmen (resp. slov délky jedna). Ta je jednoznačná, což je samotná definice slova: dvě různé posloupnosti písmen jsou různé. Dále si všimněme, že operace zřetězení je asociativní: $(a \cdot b) \cdot a = a \cdot (b \cdot a)$. Algebraicky řečeno je tedy Σ^+ s operací zřetězení *pologrupa* (struktura s jednou asociativní operací), která je *volně generovaná* slovy délky jedna (každé slovo lze jednoznačně vyjádřit pomocí písmen). Říkáme, že je to volná pologrupa nad abecedou Σ, což je její *množina volných generátoru*. Připomeňme ještě jednou, že předchozí věta je formálně nesprávná. Správně musíme říci, že Σ^+ je pologrupa slov nad abecedou Σ, která je volně generovaná slovy délky jedna. Opatrnost ve formulaci je důležitá proto, že např. množina M všech slov nad abecedou $\{a, b\}$ začínajících na a také tvoří volnou pologrupu, která je ovšem nekonečně generovaná, její volnou bazí je množina $B = \{ab^i \mid i \in \mathbb{N}\}$. Tato pologrupa je tedy isomorfní volné pologrupě B^+ nad abecedou B, je ale rozdíl mezi posloupností $(a, b, a) \in \{a, b\}^+$ a posloupností $(ab, a) \in B^+$. Množinu M, chápanou jako množina slov nad $\{a, b\}$ budeme proto raději značit $\langle B \rangle$.

> Poznamenejme, že v teorii automatů a regulárních výrazů se Kleeneovo plus používá právě ve smyslu generování množiny a mezi B^+ a $\langle B \rangle$ se nerozlišuje. To je umožněno tím, že „abecedou“ se (často bez větší reflexe) rozumí množina generátorů. Důsledkem takového přístupu ovšem je, že formálně není možné chápat slova jako posloupnosti písmen.

Prázdné slovo, které budeme značit ε (někdy se v literatuře používá λ), funguje jako jednička, tedy neutrální prvek zřetězení. Pologrupa s jednotkou se nazývá *monoid* a Σ^* je *volný monoid nad* Σ. Všimněme si, že prázdné slovo není písmeno. Vztahy jako $a \cdot \varepsilon \cdot b = \varepsilon \cdot a \cdot b$ tedy nejsou v rozporu s volností (jednoznačností rozkladu), stejně jako s ní nejsou v rozporu vztahy asociativity. Prázdné slovo způsobuje formulační nepohodlí, protože je obvykle nutné ho zmínit zvlášť (viz např. tvrzení, že délka zřetězení dvou slov je delší než libovolný z faktorů). Často budeme proto pracovat pouze v Σ^+ a jednoduché modifikace nutné pro Σ^* nebudeme výslovně zmiňovat.