Prerequisites

- general(abbreviations[w.l.o.g.])
- sets(notation, operations)
- strings(notation)

Proof: Equivalence of positive and negative grammars

This section defines both negative and positive versions of n-gram grammars and shows that they are expressively equivalent. Like in the section on the equivalence of mixed and fixed n-gram grammars, this is accomplished by a **constructive** proof. A proof is constructive if it doesn't just derive the existence of some object, but gives a concrete procedure for constructing this object. In the case at hand, the proof shows how to construct a positive grammar from a negative one, and the other way around.

Definition 1. Let Σ be some alphabet, and $\Sigma_\$$ its extension with a distinguished edge marker symbol \$. An n-gram over $\Sigma_\$$ is an element of $\Sigma_\$^n$ $(n \geq 1)$. An n-gram grammar G over alphabet Σ is a finite set of n-grams over $\Sigma_\$$. Every n-gram grammar has a **polarity**:

1. If G is negative (^-G), then a string s over Σ is well-formed with respect to ^-G iff there are no u, v over $\Sigma_\$$ and no $g \in^- G$ such that $\$^{n-1} \cdot s \cdot \$^{n-1} = u \cdot g \cdot v$.

2. If G is positive (^+G), then a string s over Σ is well-formed with respect to ^+G iff for all u, v over $\Sigma_\$$ and $g \in \Sigma_\$^n$ such that $\$^{n-1} \cdot s \cdot \$^{n-1} = u \cdot g \cdot v$, it holds that $g \in^+ G$.

The **language of** G, denoted $L(G)$, contains all strings that are well-formed with respect to G, and only those.

Theorem 2. For every n-gram grammar G there exists a grammar G' of opposite polarity such that $L(G) = L(G')$. $\lrcorner$

Proof. We assume w.l.o.g. that G is a positive grammar and denote it by ^+G. We define a negative counterpart $^-G'$ as $\Sigma_\$^n -^+ G$ and show that $L(^+G) = L(^-G')$.

First, every $s \in L(^+G)$ is also a member of $L(^-G')$. Assume towards a contradiction that $s \notin L(^-G')$. Then there must be some $g \in^- G'$ such that $\$^{n-1}s\$^{n-1} = u \cdot g \cdot v$ $(u, v, \in \Sigma_\$^*)$. But since $^-G' := \Sigma_\$^n -^+ G$, it must hold that $g \notin^+ G$, wherefore $s \notin L(^+G)$. As this contradicts our initial assumption that $s \in L(^+G)$, it cannot be the case that $s \notin L(^-G')$. So $s \in L(^-G')$ after all.

In the other direction, suppose that $s \notin L(^+G)$. Then by definition there are $u, v \in \Sigma_\$^*$ and $\Sigma_\$^n \ni g \notin^+ G$ such that $\$^{n-1}s\$^{n-1} = u \cdot g \cdot v$. But then $g \in^- G'$, which entails $s \notin L(^-G')$.

This shows that $s \in L(^+G)$ iff $s \in L(^-G')$. As s was arbitrary, this holds for all strings and establishes $L(^+G) = L(^-G')$, which concludes our proof. $\square$